High-Dimensional Statistical Inference: Phase Transition, Power Enhancement, and Sampling

by

Yinqiu He

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2021

Doctoral Committee:

Associate Professor Gongjun Xu, Co-Chair
Professor Xuming He, Co-Chair
Assistant Professor Yuekai Sun
Professor Peter Xuekun Song

Yinqiu He

yqhe@umich.edu

ORCID iD: 0000-0002-4829-033X

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my Ph.D. advisors Dr. Gongjun Xu and Dr. Xuming He, who are amazing advisors that one could ever hope for. This dissertation would not have been possible without their continuous guidance, support, and encouragement during my graduate study. Gongjun is an exemplary scholar, who is always hardworking and dedicated to research; an inspirational advisor, who is incredibly generous with his ideas and time; and a caring mentor, who constantly offers his support and empathy. I am deeply indebted to Gongjun for devoting so much time and energy to mentoring me and guiding me through the transition from a student to a researcher. His passion for research and dedication to mentoring have profoundly influenced me from both professional and personal perspectives. I am also fortunate enough to work with Xuming. During our interactions, I am constantly amazed by his sharpness on research, his ability of distilling the essence of a problem rapidly, and his broad knowledge and deep insights about many different topics. Our discussions have been a great source of inspiration and greatly shaped my way of approaching research problems. In retrospect, having Gongjun and Xuming as my advisors was one of the best things that could ever happen to me.

I am also very grateful to Dr. Peter Xuekun Song and Dr. Yuekai Sun for serving on my doctoral dissertation committee and providing me invaluable comments. I first met Peter through taking his course BIOSTAT 885. Later, I was fortunate to collaborate with him on mediation analysis and attend his group meetings. Peter's enthusiasm for research has left a lot of positive impacts on me. I would also like to thank Yuekai for having interesting discussions on research problems with me. His humor makes those meetings enjoyable.

My thanks also go to other faculty members and staff members at the University of University who have helped me over the past few years. I would like to thank Liza Levina for sharing her experiences and giving support during the Women in Statistics lunch series and other interactions. I am very grateful to Ji Zhu for sharing a lot of comments and insights with me in our discussions. I am also thankful to Naisyin Wang for her strong encouragement and support. I would like to thank Xu Shi and Kean Ming Tan for providing helpful suggestions on my job talk. I want to thank all the staff members, including Judy, Jean, Bebe, and Gina, who always patiently helped with my questions and warmly welcomed me into the office with big smiles. I was also fortunate to collaborate with a number of wonderful researchers outside Michigan, including Wei Pan, Chong Wu, and Tiefeng Jiang. I would like to thank them for their help in terms of both research and the job search.

To all my friends that I made throughout the Ph.D. studies: This journey would not have been so enjoyable without you. I especially thank everyone in our research groups, in our office, and in our cohort. It was a great fun to spend time with you, and I have learned a lot from our conversations. I also want to thank my housemates, whose companion has made my graduate life so delightful. Last but not least, I would like to thank my parents and grandparents for their unwavering support.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

ABSTRACT

The "Big Data" era features large amounts of high-dimensional data, in which the number of characteristics per subject is large. The high dimensionality of such big data can pose many new challenges for statistical inference, including (I) the invalidity of classical approximation theory, (II) the loss of statistical power, and (III) the increase of computational burden. This dissertation studies three important problems that arise in this context.

(I) The first part introduces a newly discovered phase transition phenomenon of the widely used likelihood ratio tests. In particular, it is broadly recognized that classical large-sample approximation theory that is valid under finite dimensions may fail under high dimensions. But there is usually a lack of understanding of when such transition happens as the data dimension increases. This issue can hinder the validation of statistical inference in practice. Focusing on the popular likelihood ratio tests, we derive necessary and sufficient conditions characterizing the phase transition boundaries where Wilks' theorem becomes invalid. Based on this, we further obtain sharp characterization of the approximation bias of Wilks' theorem.

(II) The second part proposes a novel adaptive testing framework that can maintain high statistical power against a variety of alternative hypotheses. Particularly, many scientific questions in high-dimensional data analyses can be formulated as testing high-dimensional parameters globally, e.g., testing whether there exists any association between a large number of SPNs and certain heritable disease in genome-wide association studies. In these problems, many existing methods are designed to

capture certain directional information in a high-dimensional space and thus only powerful for specific alternatives. To enhance the statistical power, we construct an innovative family of test statistics that can capture the information in different directions of a high-dimensional space. For a broad class of problems, we establish high-dimensional asymptotic theory for the constructed statistics and develop testing procedures that are adaptively powerful across a wide range of scenarios.

(III) The third part concerns the computational challenge of quantifying rare-event probabilities in statistical inference. In particular, analyzing high-dimensional data frequently involves a large number of hypotheses and results in stringent significance thresholds. It is therefore often required to accurately estimate an extreme tail probability of each test statistic. However, analytical formulae are usually unavailable for nontrivial statistics, and naive Monte Carlo methods usually require a huge number of simulations and are computationally costly. Driven by rare-event issues arising from testing covariance structures, we develop an asymptotically efficient importance sampling algorithm to compute the extreme tail probabilities of the popular ratio statistic of the largest eigenvalue to the trace of a Wishart matrix.

CHAPTER I

Introduction

Rapid developments of high-throughput biomedical technologies produce large amounts of massive and high-dimensional data (Marx, 2013; Chattopadhyay and Lu, 2019; Palit et al., 2019). Such data often contain a large number of measurements or features that can be great resources of information in a variety of scientific studies. For instance, Alzheimer's Disease Neuroimaging Initiative (ADNI; `http://adni.loni.usc.edu`) collected extensive features, including genetics, brain images, and many other health-related indicators, which may be analyzed to understand the genetic susceptibility and progression of Alzheimer's Disease (Apostolova et al., 2018). Another example, the Trans-Omics for Precision Medicine (TOPMed) Program (`https://www.nhlbiwgs.org`) collected different types of features such as whole-genome sequencing and various omics data (e.g., metabolic profiles, epigenomics, protein and RNA expression patterns) that can be used as important scientific resources for advancing precision medicine (Taliun et al., 2021). In these large-scale data, although numerous features could provide rich information for research, the high dimensionality of data can pose new challenges for statistical inference in many ways. Examples include, but are not limited to, the following aspects.

(I) *Theoretical validity*: Approximation theory in the classical statistical inference literature (Casella and Berger, 2002) commonly assumed that the data dimension was finite compared to the sample size. Such assumptions can be violated when analyzing high-dimensional data. It follows that the classical theory may become inappropriate or inaccurate due to the effects of high dimensions (Bai and Saranadasa, 1996; Chen et al., 2009).

(II) *Statistical power*: High-dimensional data analysis often involves a large number of parameters to investigate. Different values of the parameters correspond to different scenarios of the underlying truth. Existing procedures that are designed to be statistically powerful under certain scenarios can be of limited statistical power under other scenarios (Basu and Pan, 2011). Using procedures of inadequate power might result in missed opportunities of significant scientific findings and undermine the reliability of research (Button et al., 2013b; Dumas-Mallet et al., 2017).

(III) *Computational burden*: Big data in various scientific areas are of increasing scale and complexity (Huang et al., 2015; Qin et al., 2015). Analyzing those big data with inefficient algorithms can be computationally costly or even practically infeasible (Wang et al., 2016).

This dissertation studies three problems motivated by the above challenges arising in the high-dimensional statistical inference.

(I) Likelihood Ratio Tests and Wilks' Theorem: Valid or Not? Hypothesis testing, along with the closely related concept of the confidence region, plays a crucial role in statistical inference (Lehmann, 2012). It provides a foundation for investigating the underlying scientific mechanisms and drawing conclusions from data in many applications. One fundamental and standard method for many hypothesis testing problems is the likelihood ratio test (Muirhead, 2009). In the classical settings, likelihood ratio tests have been shown to enjoy certain optimality properties (Neyman and Pearson, 1933), and Wilks' theorem offers universal chi-squared approximations for the likelihood ratio test statistics (Wilks, 1932). However, for modern datasets with increasing dimensions, researchers have found that the conventional approximations based on Wilks' theorem of the likelihood ratio test statistic often becomes inaccurate. Although new approximations have been proposed in high-dimensional settings, it is less understood when the transitions from the conventional chi-squared approximations to the new approximations happen.

To address this issue, Chapter II studies an interesting new discovery of phase transition phenomena of Wilks' theorem in a diverse range of problems, including the tests of multivariate means and covariances, the exploratory factor analysis, and the multivariate linear regression. In each setting, we derive the *necessary* and *sufficient* condition characterizing the phase transition boundary of Wilks' theorem. The condition is specified through the increasing rates of the data dimension with respect to the sample size. Under the asymptotic regime when the condition is satisfied, Wilks' theorem holds; when the condition is violated, Wilks' theorem fails and alternative high-dimensional approximations should be used. Based on the phase transition conditions, we further derive the asymptotic biases of the chi-squared approximations. Under the asymptotic regime when Wilks' theorem holds, the derived asymptotic bias sharply characterizes the convergence rate of the distribution of the likelihood ratio test statistic to the limiting chi-squared distribution. When Wilks' theorem fails, the derived asymptotic bias describes the unignorable discrepancy between the chi-squared approximation and the true distribution of the likelihood ratio test statistic. These results could provide helpful insights into the use of the chi-squared approximations in scientific practices. The materials in this chapter contribute to research papers He, Meng, Zeng, and Xu (2021b), He, Jiang, Wen, and Xu (2021a) and He, Wang, and Xu (2021c).

(II) High-Dimensional Testing: How to Enhance Statistical Power? When the validity of statistical inference procedures can be guaranteed, the next key statistical inquiry is to enhance the statistical power, which could increase the chance for true scientific discoveries and improve the reliability of research (Button et al., 2013a; Dumas-Mallet et al., 2017). In particular, consider the questions in high-dimensional data analyses that can be formulated as globally testing the overall patterns of high-dimensional parameters. For example, in genome-wide association studies, it can be of interest to test whether there exists any nonzero association between a large number of single nucleotide polymorphisms (SPNs) in a genetic marker set and certain heritable disease (Wang et al., 2011; Kim et al., 2016). In these problems, high-dimensional parameters of interest (e.g., associations between numerous SNPs and a disease) can induce a large parameter space. But many existing tests are designed to capture certain directional information of the large space and thus only powerful for specific alternative hypotheses (Basu and Pan, 2011; Kim et al., 2014) . Since the underlying truth is usually unknown, it can be unclear how to choose a powerful test in practice. Using procedures of low statistical power could lead to the missed opportunities of important scientific findings in applications (Sham and Purcell, 2014).

To conduct powerful tests against a variety of alternatives, Chapter III develops a framework that can achieve high power adaptively in a large class of global hypothesis testing problems, including tests of multivariate means, covariances, and coefficients in generalized linear models. Specifically, we construct an innovative family of U-statistics as test statistics. Through the mindful construction, U-statistics in the family share the desirable property of unbiasedness for hypothesis testing and can cover the statistics in other popular methods as special cases. We further establish new high-dimensional theory for the family of U-statistics and show an interesting phenomenon of asymptotic independence among U-statistics of different orders. This phenomenon is a new theoretical finding and suggests that different U-statistics in the family can intuitively capture the information in different orthogonal directions in a high-dimensional space. The nice theoretical properties enable us to develop an adaptive testing procedure that combine different U-statistics in the family together. It follows that framework covers many popular methods that only focus on one or two particular alternatives as special cases and consequently achieves high power across different scenarios. This is particularly useful when the underlying truth is unknown in practice. We also develop a computationally efficient algorithm to compute the proposed U-statistics. The materials in this chapter contribute to the research paper He, Xu, Wu, and Pan (2021d).

(III) Rare-Event Probabilities: How to Compute Efficiently? The large scale of data can lead to various computational challenges in statistical inference (Wang et al., 2016). In this dissertation, Chapter IV considers one problem on the efficient computation of rare-event probabilities (Bucklew, 2013). Particularly, analyzing high-dimensional data frequently involves a large number of hypotheses and results in stringent significance thresholds (Xu and Wang, 2020). It is thus required to accurately estimate an extreme tail probability of each test statistic. Estimating tail probabilities in a very small magnitude is often challenging, since analytical formulae are usually unavailable for nontrivial statistics due to the non-standard extreme tail behaviors, and naive Monte Carlo methods usually require a huge number of simulations and thus are computationally costly (Asmussen and Glynn, 2007).

In particular, Chapter IV studies the ratio of the largest eigenvalue to the trace of a white Wishart matrix, which plays an important role on scale-independent testing of covariance structure (Anderson, 2003; Muirhead, 2009). Despite its importance, there is no simple-to-compute expression for the exact distribution of the ratio statistic. Our research proposes an importance sampling algorithm for estimating the tail probability of the ratio statistic. We utilize a sampling measure that approximates the conditional distribution of the ratio statistic. We also give a theoretical analysis based on a large deviation result in random matrix theory, which shows that the algorithm is asymptotically efficient. In addition, simulation studies show that it is computationally more efficient than the naive Monte Carlo methods and outperforms existing approaches based on asymptotic approximations, especially when estimating probabilities of rare events. The materials in this chapter contribute to the research paper He and Xu (2018).

Notation We introduce some notation to be used in the rest of this dissertation. For two series of numbers $u_{n,p}$ and $v_{n,p}$ that change with n and p: $u_{n,p} = o(v_{n,p})$ denotes $\limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| = 0$; $u_{n,p} = O(v_{n,p})$ denotes $\limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| < \infty$; $u_{n,p} = \Theta(v_{n,p})$ denotes $0 < \liminf_{n,p\to\infty} |u_{n,p}/v_{n,p}| \leq \limsup_{n,p\to\infty} |u_{n,p}/v_{n,p}| < \infty$; $u_{n,p} \simeq v_{n,p}$ denotes $\lim_{n,p\to\infty} u_{n,p}/v_{n,p} = 1$. Moreover, $\xrightarrow{P}$ and $\xrightarrow{D}$ represent the convergence in probability and distribution respectively.

CHAPTER II

Phase Transition Phenomena of Likelihood Ratio Tests

2.1 Introduction

The likelihood ratio test is a standard testing method for many hypothesis testing problems due to its nice statistical properties (Anderson, 2003; Muirhead, 2009). Under low-dimensional settings, classic theorems offer general asymptotic results for various likelihood ratio test statistics. One of the most celebrated and fundamental results is Wilks' theorem, which states that, under the null hypothesis, twice the negative log-likelihood ratio asymptotically approaches a χ^2_f distribution, where f is the difference of the degrees of freedom between the null and alternative hypotheses. The popularly used Bartlett correction provides a general rescaling strategy that further improves the finite sample accuracy of the chi-squared approximations (Cordeiro and Cribari-Neto, 2014; Barndorff-Nielsen and Hall, 1988). Similar Wilks' phenomenon and Bartlett correction were also studied for the empirical likelihood (Owen, 1990; DiCiccio et al., 1991; Chen and Cui, 2006).

Despite the extensive literature on the Wilks'-type phenomenon of likelihood ratio tests under finite dimensions, it is of emerging interest to study the asymptotic regimes with the large sample size n and the diverging data dimension p in a wide variety of modern applications. To understand how large the dimension p can be to ensure the validity of the classical Wilks' phenomenon, various works establish sufficient conditions on the growth rate of p as n increases. For instance, Portnoy (1988) showed that the chi-squared approximation of the likelihood ratio test statistic for a simple hypothesis in canonical exponential families holds if $p/n^{2/3} \to 0$. Moreover, Hjort et al. (2009), Chen et al. (2009), and Tang and Leng (2010) studied the empirical likelihood ratio statistic when $p \to \infty$. Particularly, Chen et al. (2009) argued that $p/n^{1/2} \to 0$ is likely to be the best rate for the chi-squared approximation of general empirical likelihood ratio test, and showed that for the least-squares empirical likelihood, a simplified version of the empirical likelihood, the chi-squared approximation holds if $p/n^{2/3} \to 0$. The effect of data dimension was also studied in other inference problems; see, for example, Portnoy (1985), He and Shao (2000), and Wang (2011).

When the dimension p further increases, researchers have found that the chi-squared approximations based on Wilks' theorem often become inaccurate, resulting in the failure of the corresponding likelihood ratio tests. To address this issue, various corrections and alternative approximations for the likelihood ratio tests have been proposed. For example, when p is asymptotically proportional to n, namely, $p/n \to y \in (0, 1)$ as $n \to \infty$, Bai et al. (2009), Jiang and Yang (2013), and Jiang and Qi (2015) proposed normal approximations for the corrected likelihood ratio tests on mean vectors and covariance matrices. Zheng (2012), Bai et al. (2013), and He et al. (2021a) proposed normal approximations for corrected likelihood ratio tests in multivariate linear regression models. Furthermore, Sur and Candès (2019), Sur et al. (2019), and Candès and Sur (2020) studied the phase transition of the maximum likelihood estimator for the logistic regression and proposed a rescaled chi-squared approximation for the likelihood ratio test.

Despite the proposed distributional theory of the likelihood ratio tests for low- or high-dimensional data, there still lacks a quantitative guideline on which approximation should be chosen to use in practice, especially for moderate-dimensional data. For instance, when analyzing a dataset with the number of parameters $p \leq 5$ and sample size $n = 100$, the chi-squared approximation may be considered as reliable. However, when studying a data set with moderate dimension, e.g., p is between 6 to 20 and sample size $n = 100$, it may be unclear to practitioners whether they can still apply the classical chi-squared approximations or they should turn to other high-dimensional asymptotic results. To address this practical issue, it is of interest to investigate the phase transition boundary where the chi-squared approximation starts to fail as p increases, and also characterize the approximation accuracy. Theoretically, this needs a deep understanding of the limiting behavior of the likelihood ratio test statistics from low to high dimensions.

The remaining part of this chapter studies the fundamental phase transition phenomenon of Wilks' theorem in three important classes of testing problems. In particular, Section 2.2 starts with several standard likelihood ratio tests of multivariate mean vectors and covariance matrices. Section 2.3 discusses an extension to exploratory factor analysis. Section 2.4 further considers the multivariate linear regression model, where both the dimensions of the predictors and the responses need to be investigated.

2.2 Results for Multivariate Means and Covariances

In this section, we focus on several standard likelihood ratio tests on multivariate mean and covariance structures that are widely used in biomedical and social sciences (Pituch and Stevens, 2015; Cleff, 2019). For each considered likelihood ratio test, we derive its phase transition boundary of Wilks' phenomenon and also provide an in-depth analysis of the accuracy of the chi-squared approximation. First, in terms of the phase transition boundary, we establish the *necessary* and *sufficient* condition for Wilks' theorem to hold when p increases with n. Specifically, we show that the chi-squared approximations hold if and only if $p/n^d \to 0$, where the value of d depends on the testing problem and whether the Bartlett correction is used. Interestingly, the proposed phase transition boundaries resonate with the abovementioned literature (e.g., Portnoy, 1988; Chen et al., 2009), which mostly focused on sufficient conditions without the Bartlett correction. Second, we provide a detailed characterization of the asymptotic bias of each chi-squared approximation. Specifically, we consider two local asymptotic regimes, depending on whether Wilks' theorem holds or not. Under the asymptotic regime when Wilks' theorem holds, the derived asymptotic bias sharply characterizes the convergence rate of the distribution of the likelihood ratio test statistic to the limiting chi-squared distribution, and thus provides a useful measure on the accuracy of the chi-squared approximation. When Wilks' theorem fails, the derived asymptotic bias describes the unignorable discrepancy between the chi-squared approximation and the true distribution of the likelihood ratio test statistic. As illustrated in the simulation studies, our theoretical results of the phase transition boundaries and the asymptotic biases may provide a helpful guideline on the use of the chi-squared approximations in practice. In the following, Sections 2.2.1 and 2.2.2 present the theoretical results of one-sample and two-sample tests, respectively. Section 2.2.3 include simulation studies. All the technical proofs are included in Appendix A.

2.2.1 One-Sample Tests

Under one-sample problems, suppose $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ are independent and identically distributed random vectors with distribution $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which denotes a p-variate multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We define $\overline{\mathbf{x}} = n^{-1}\sum_{i=1}^n \mathbf{x}_i$ and $\mathbf{A} = \sum_{i=1}^n (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\intercal$, and denote the determinant and the trace of $\mathbf{A}$ by $|\mathbf{A}|$ and $\mathrm{tr}(\mathbf{A})$, respectively. We next introduce the considered testing problems and the corresponding likelihood ratio tests (Anderson, 2003; Muirhead, 2009).

(I) *Testing Specified Value for the Mean Vector.* This test examines whether the population mean vector $\boldsymbol{\mu}$ is equal to a specified vector $\boldsymbol{\mu}_0 \in \mathbb{R}^p$, that is, $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_a : H_0$ is not true. Through the transformation $\mathbf{x}_i - \boldsymbol{\mu}_0$, we consider, without loss of generality, $\boldsymbol{\mu}_0 = (0, \ldots, 0)^\intercal$. Then, the likelihood ratio test statistic is $\Lambda_n = |\mathbf{A}|^{n/2}(\mathbf{A} + n\bar{\mathbf{x}}\bar{\mathbf{x}}^\intercal)^{-n/2}$. When p is fixed and $n \to \infty$, under the null hypothesis, the classical chi-squared approximation without correction is $-2\log\Lambda_n \xrightarrow{d} \chi^2_f$, where $\xrightarrow{d}$ represents the convergence in distribution and $f = p$, and the chi-squared approximation with the Bartlett correction is $-2\rho\log\Lambda_n \xrightarrow{d} \chi^2_f$, where $\rho = 1 - (1 + p/2)/n$.

(II) *Testing the Sphericity of the Covariance Matrix.* This test examines whether the covariance matrix $\boldsymbol{\Sigma}$ is proportional to an identity matrix; that is, $H_0 : \boldsymbol{\Sigma} = \lambda \mathrm{I}_p$ against $H_a : H_0$ is not true, where $\lambda > 0$ is an unspecified constant and I_p denotes the $p \times p$ identity matrix. The likelihood ratio test statistic is $\Lambda_n = |\mathbf{A}|^{(n-1)/2} \{\mathrm{tr}(\mathbf{A})/p\}^{-p(n-1)/2}$. When p is fixed and $n \to \infty$, under the null hypothesis, the chi-squared approximation is $-2\log\Lambda_n \xrightarrow{d} \chi^2_f$, where $f = (p-1)(p+2)/2$, and the chi-squared approximation with the Bartlett correction is $-2\rho\log\Lambda_n \xrightarrow{d} \chi^2_f$, where $\rho = 1 - \{6(n-1)p\}^{-1}(2p^2 + p + 2)$.

(III) *Joint Testing Specified Values for the Mean Vector and Covariance Matrix.* Consider a specified vector $\boldsymbol{\mu}_0 \in \mathbb{R}^p$ and a specified positive-definite matrix $\boldsymbol{\Sigma}_0 \in \mathbb{R}^{p\times p}$. We study the test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ against $H_a : H_0$ is not true. By applying the transformation $\boldsymbol{\Sigma}_0^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu}_0)$, we assume, without loss of generality, that $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\boldsymbol{\Sigma}_0 = \mathrm{I}_p$. Then, the likelihood ratio test statistic is $\Lambda_n = (e/n)^{np/2} |\mathbf{A}|^{n/2} \exp\{-\mathrm{tr}(\mathbf{A})/2 - n\bar{\mathbf{x}}^\intercal\bar{\mathbf{x}}/2\}$. When p is fixed and $n \to \infty$, under the null hypothesis, the chi-squared approximation is $-2\log\Lambda_n \xrightarrow{d} \chi^2_f$, where $f = p(p+3)/2$, and the chi-squared approximation with the Bartlett correction is $-2\rho\log\Lambda_n \xrightarrow{d} \chi^2_f$, where $\rho = 1 - \{6n(p+3)\}^{-1}(2p^2 + 9p + 11)$.

For the likelihood ratio tests of the above three testing problems, Theorem 2.2.1 gives the phase transition boundaries of the chi-squared approximations without and with the Bartlett correction.

Theorem 2.2.1 (Phase Transition Boundaries)**.** *Assume $n > p+1$ and $n-p \to \infty$ as $n \to \infty$. Under H_0, for the chi-squared approximations without and with the Bartlett correction of each likelihood ratio test in (I)–(III), we have the following necessary and sufficient conditions:*

(i) $\sup_{\alpha\in(0,1)} |\Pr\{-2\log\Lambda_n > \chi^2_f(\alpha)\} - \alpha| \to 0$ *if and only if* $p/n^{d_1} \to 0$;

(ii) $\sup_{\alpha\in(0,1)} |\Pr\{-2\rho\log\Lambda_n > \chi^2_f(\alpha)\} - \alpha| \to 0$ *if and only if* $p/n^{d_2} \to 0$,

where the values of d_1 and d_2 under the three testing problems are listed in the table below.

	(I) Mean	*(II) Covariance*	*(III) Joint*
(i) without correction d_1:	2/3	1/2	1/2
(ii) with correction d_2:	4/5	2/3	2/3

In Theorem 2.2.1, $n > p + 1$ is assumed to ensure the existence of the likelihood ratio tests. We next discuss the obtained phase transition boundaries of the classical chi-squared approximations without correction. When only testing mean parameters, Theorem 2.2.1 suggests that the chi-squared approximation holds if and only if $p/n^{2/3} \to 0$. This asymptotic regime is similarly assumed in Portnoy (1988), which considered testing p natural parameters in exponential families. However, Portnoy (1988) only showed the sufficiency of $p/n^{2/3} \to 0$ for the chi-squared approximation to be applied, and did not establish the necessary and sufficient result, which is essential for understanding the phase transition behaviors. In addition, when the likelihood ratio tests involve covariance matrices as in (II) and (III), Theorem 2.2.1 shows that the chi-squared approximation holds if and only if $p/n^{1/2} \to 0$, which is consistent with the discussion in Chen et al. (2009). Particularly, under certain regularity conditions, Chen et al. (2009) established that the chi-squared approximation of the empirical likelihood ratio test holds if $p/n^{1/2} \to 0$. The authors further argued that $p/n^{1/2} \to 0$ is likely to be the best rate for p, because it is the necessary and sufficient condition for the convergence of the sample covariance matrix to the true covariance matrix $\mathbf{\Sigma}$ under the trace norm when the eigenvalues of $\mathbf{\Sigma}$ are bounded. The analysis provides an intuitive explanation for the phase transition boundaries obtained above, and our necessary and sufficient result would serve as another support for their conjecture, despite the different problem settings in Chen et al. (2009) and here.

Additionally, for the chi-squared approximations with the Bartlett correction, Theorem 2.2.1 also explicitly characterizes their phase transition boundaries, which generally achieve a larger asymptotic region than those without correction. When p is fixed, the Bartlett correction serves as a rescaling strategy that can improve the convergence rate of the likelihood ratio statistic from $O(n^{-1})$ to $O(n^{-2})$; however, when p grows with sample size n, the classical result cannot apply directly. Alternatively, the results in Theorem 2.2.1 provide a precise illustration of how the Bartlett correction improves the chi-squared approximations in terms of the phase transition boundaries.

The phase transition boundaries in Theorem 2.2.1 give the necessary and sufficient conditions on the asymptotic regimes of (n, p) in Wilks' phenomenon. When applying the likelihood ratio test in practice, it is desired to have a better understanding of the accuracy of the chi-squared approximation, especially near its phase transition boundary. The following Theorem 2.2.2 characterizes the accuracy of each chi-squared approximation for tests (I)–(III) when Wilks' phenomenon holds. Specifically, we consider the asymptotic regime where (n, p) satisfies the corresponding necessary and sufficient condition in Theorem 2.2.1, i.e., $p/n^{d_1} \to 0$ and $p/n^{d_2} \to 0$ for the chi-squared approximations without and with the Bartlett correction, respectively.

Theorem 2.2.2 (Asymptotic Biases)**.** *For each likelihood ratio test (I)–(III), let d_i, $i = 1, 2$ take the corresponding values in Theorem 2.2.1. Let z_α denote the upper α-level quantile of the standard normal distribution. Consider $p \to \infty$ as $n \to \infty$. Then under H_0, given $\alpha \in (0,1)$,*

(i) when $p/n^{d_1} \to 0$, the chi-squared approximation satisfies

$$\Pr\{-2\log\Lambda_n > \chi_f^2(\alpha)\} - \alpha = \frac{\vartheta_1(n,p)}{\sqrt{\pi}} \exp\Big(-\frac{z_\alpha^2}{2}\Big) + o\Big(\frac{p^{1/d_1}}{n}\Big); \qquad (2.1)$$

(ii) when $p/n^{d_2} \to 0$, the chi-squared approximation with the Bartlett correction satisfies

$$\Pr\{-2\rho\log\Lambda_n > \chi_f^2(\alpha)\} - \alpha = \frac{\vartheta_2(n,p)}{\sqrt{\pi}} \exp\Big(-\frac{z_\alpha^2}{2}\Big) + o\Big(\frac{p^{2/d_2}}{n^2}\Big). \qquad (2.2)$$

The values of $\vartheta_1(n,p)$ and $\vartheta_2(n,p)$ under three testing problems (I)–(III) are listed below.

(I) Mean:

$$\vartheta_1(n,p) = \frac{p^2+2p}{4n\sqrt{f}},$$

$$\vartheta_2(n,p) = \frac{p(p^2-4)}{24(\rho n)^2\sqrt{f}};$$

(II) Covariance:

$$\vartheta_1(n,p) = \frac{p(2p^2+3p-1)-4/p}{24(n-1)\sqrt{f}},$$

$$\vartheta_2(n,p) = \frac{(p-2)(p-1)(p+2)(2p^3+6p^2+3p+2)}{144p^2\rho^2(n-1)^2\sqrt{f}};$$

(III) Joint:

$$\vartheta_1(n,p) = \frac{p\,(2p^2+9p+11)}{24n\sqrt{f}},$$

$$\vartheta_2(n,p) = \frac{p(2p^4+18p^3+49p^2+36p-13)}{144(p+3)(\rho n)^2\sqrt{f}}.$$

In Theorem 2.2.2, the forms of $\vartheta_1(n,p)$ and $\vartheta_2(n,p)$ are derived from a nontrivial calculation of certain complicated infinite series (see Eq. (A.20) and (A.28) in the

Appendix). We can see that for each test, $\vartheta_1(n,p)$ and $\vartheta_2(n,p)$ are of orders of $p^{1/d_1}n^{-1}$ and $p^{2/d_2}n^{-2}$, respectively.

In the above discussion, we focus on the local asymptotic regime when Wilks' phenomenon holds, and the derived bias describes the accuracy of the chi-squared approximation. When p further increases beyond this local asymptotic regime, the chi-squared approximation starts to fail, and the approximation bias becomes asymptotically unignorable. The following Theorem 2.2.3 characterizes such unignorable biases of the chi-squared approximations. Particularly, we consider the local asymptotic regime $p/n \to 0$, which includes the case when Wilks' theorem fails, that is, $p/n^{d_1} \not\to 0$ for the chi-squared approximation, and $p/n^{d_2} \not\to 0$ for the chi-squared approximation with the Bartlett correction.

Theorem 2.2.3 (Asymptotic Biases)**.** *Assume $p \to \infty$ and $p/n \to 0$ as $n \to \infty$. For each likelihood ratio test (I)–(III), under H_0, there exists a small constant $\delta \in (0,1)$ such that for any $\alpha \in (0,1)$,*

(i) the chi-squared approximation satisfies

$$\Pr\left\{-2\log\Lambda_n > \chi^2_f(\alpha)\right\} - \alpha = \bar{\Phi}\left\{\frac{\chi^2_f(\alpha) + 2\mu_n}{2n\sigma_n}\right\} - \alpha + O\left\{\left(\frac{p}{n}\right)^{\frac{1-\delta}{2}} + f^{-\frac{1-\delta}{6}}\right\}, \tag{2.3}$$

where $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$, and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution;

(ii) the chi-squared approximation with the Bartlett correction satisfies

$$\Pr\left\{-2\rho\log\Lambda_n > \chi^2_f(\alpha)\right\} - \alpha = \bar{\Phi}\left\{\frac{\chi^2_f(\alpha) + 2\rho\mu_n}{2\rho n\sigma_n}\right\} - \alpha + O\left\{\left(\frac{p}{n}\right)^{\frac{1-\delta}{2}} + f^{-\frac{1-\delta}{6}}\right\}. \tag{2.4}$$

The values of μ_n and σ_n under each problem are listed below, where $L_{x,p} = \log(1-p/x)$

for $x > p$.

(I) *Mean:*
$$\mu_n = \frac{n}{2}\Big\{\Big(n - p - \frac{3}{2}\Big)(L_{n,p} - L_{n-1,p}) + L_{n,p} + pL_{n,1}\Big\},$$
$$\sigma_n^2 = \frac{1}{2}(L_{n,p} - L_{n-1,p});$$

(II) *Covariance:*
$$\mu_n = -\frac{n-1}{2}\{(n - p - 3/2)L_{n-1,p} + p\},$$
$$\sigma_n^2 = -\frac{1}{2}\left(\frac{p}{n-1} + L_{n-1,p}\right)\frac{(n-1)^2}{n^2};$$

(III) *Joint:*
$$\mu_n = -\frac{n}{2}\{(n - p - 3/2)L_{n-1,p} + p\} - \frac{p}{2},$$
$$\sigma_n^2 = -\frac{1}{2}\left(\frac{p}{n-1} + L_{n-1,p}\right).$$

Theorem 2.2.3 is derived by quantifying the difference between the characteristic functions of $\log \Lambda_n$ and a normal distribution (see Lemma A.1.4 in the Appendix). The local asymptotic regime $p/n \to 0$ is assumed mainly for the technical simplicity of evaluating the asymptotic expansions of the characteristic functions. Under the conditions of Theorem 2.2.3, $\bar{\Phi}[\{\chi_f^2(\alpha) + 2\mu_n\}/(2n\sigma_n)] - \alpha$ in (2.3) can be approximated by $\bar{\Phi}\{z_\alpha + (f + 2\mu_n)/(2n\sigma_n)\} - \bar{\Phi}(z_\alpha)$, where $(f + 2\mu_n)/(2n\sigma_n)$ is of the order of pn^{-d_1} (see Remark A.3 in the Supplementary Material). Consequently, when the chi-squared approximation fails, i.e., $pn^{-d_1} \not\to 0$, we know that $\bar{\Phi}[\{\chi_f^2(\alpha) + 2\mu_n\}/(2n\sigma_n)] - \alpha$ in (2.3) characterizes the corresponding unignorable bias of the chi-squared approximation. Similarly, we can show that $\bar{\Phi}[\{\chi_f^2(\alpha) + 2\rho\mu_n\}/(2\rho n\sigma_n)] - \alpha$ can be approximated by $\bar{\Phi}\{z_\alpha + (f + 2\rho\mu_n)/(2\rho n\sigma_n)\} - \bar{\Phi}(z_\alpha)$, where $(f + 2\rho\mu_n)/(2\rho n\sigma_n)$ is of the order of $p^{2/d_2}n^{-2}$. Therefore, when the chi-squared approximation with the Bartlett correction fails, i.e., $pn^{-d_2} \not\to 0$, we know that (2.4) characterizes the corresponding unignorable approximation bias.

Remark II.1. *Although the above discussions consider $p/n^{d_1} \not\to 0$ and $p/n^{d_2} \not\to 0$,* (2.3) *and* (2.4) *in Theorem 2.2.3 also hold under the asymptotic regimes $p/n^{d_1} \to 0$ and $p/n^{d_2} \to 0$ examined in Theorem 2.2.2. However, since Theorems 2.2.2 and 2.2.3* *focus on different asymptotic regimes and are proved using different techniques, we can show that when* $p/n^{d_1} \to 0$ *and* $p/n^{d_2} \to 0$, (2.3) *and* (2.4) *have an additional remainder term* $O\{(p/n)^{(1-\delta)/2} + f^{-(1-\delta)/6}\}$ *compared to* (2.1) *and* (2.2)*, respectively; see Remark A.3 in the Supplementary Material. Therefore, under the asymptotic regimes of Theorem 2.2.2,* (2.1) *and* (2.2) *provide a sharper characterization of the accuracy of the chi-squared approximations than* (2.3) *and* (2.4)*, respectively.*

2.2.2 Multiple-Sample Tests

Under the multiple-sample problems, let k denote the number of samples, which is assumed to be fixed compared to the sample size. In each sample $i = 1, \ldots, k$, the observations $\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in_i}$ are independent and identically distributed $\mathcal{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ random vectors. In this subsection, we define $\overline{\mathbf{x}}_i = n_i^{-1} \sum_{j=1}^{n_i} \mathbf{x}_{ij}$ and $\mathbf{A}_i = \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)^\intercal$ for $i = 1, \ldots, k$, and let $\mathbf{A} = \mathbf{A}_1 + \ldots + \mathbf{A}_k$ and $n = n_1 + \ldots + n_k$. We next briefly review three multiple-sample likelihood ratio tests.

(IV) *Testing the Equality of Several Mean Vectors.* Consider $H_0 : \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_k$ agains $H_a : H_0$ is not true, where the covariances of the k samples are assumed to be the same. Define $\mathbf{B} = \sum_{i=1}^{k} n_i (\overline{\mathbf{x}}_i - \overline{\mathbf{x}})(\overline{\mathbf{x}}_i - \overline{\mathbf{x}})^\intercal$ and $\overline{\mathbf{x}} = n^{-1} \sum_{i=1}^{k} n_i \overline{\mathbf{x}}_i$. Then, the likelihood ratio test statistic is $\Lambda_n = |\mathbf{A}|^{n/2} |\mathbf{A} + \mathbf{B}|^{-n/2}$. When p is fixed and $n \to \infty$, the chi-squared approximation is $-2 \log \Lambda_n \xrightarrow{d} \chi^2_f$, where $f = (k-1)p$, and the chi-squared approximation with the Bartlett correction is $-2\rho \log \Lambda_n \xrightarrow{d} \chi^2_f$, where $\rho = 1 - \{1 + (k+p)/2\}/n$.

(V) *Testing the Equality of Several Covariance Matrices.* Consider $H_0 : \boldsymbol{\Sigma}_1 = \ldots = \boldsymbol{\Sigma}_k$ against $H_a : H_0$ is not true. For this test, $\Lambda_n = |\mathbf{A}|^{-(n-k)/2} (n-k)^{(n-k)p/2} \times \prod_{i=1}^{k} (n_i - 1)^{-(n_i-1)p/2} |\mathbf{A}_i|^{(n_i-1)/2}$ is the modified likelihood ratio test statistic with the unbiasedness property. When p is fixed and $\min_{1 \le i \le k} n_i \to \infty$, the chi-squared approximation is $-2 \log \Lambda_n \xrightarrow{d} \chi^2_f$, where $f = p(p+1)(k-1)/2$, and the chi-squared approximation with the Bartlett correction is $-2\rho\log\Lambda_n \xrightarrow{d} \chi^2_f$, where $\rho = 1 - \{6(p+1)(k-1)\}^{-1}(2p^2+3p-1)\{\sum_{i=1}^k (n_i-1)^{-1} - (n-k)^{-1}\}$.

(VI) *Joint Testing the Equality of Mean Vectors and Covariance Matrices.* Consider $H_0 : \boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_k,\ \boldsymbol{\Sigma}_1 = \ldots = \boldsymbol{\Sigma}_k$ against $H_a : H_0$ is not true. The likelihood ratio test statistic is $\Lambda_n = n^{pn/2}|\mathbf{A}+\mathbf{B}|^{-n/2} \times \prod_{i=1}^k n_i^{-pn_i/2}|\mathbf{A}_i|^{n_i/2}$. When p is fixed and $\min_{1\le i\le k} n_i \to \infty$, the chi-squared approximation is $-2\log\Lambda_n \xrightarrow{d} \chi^2_f$, where $f = p(k-1)(p+3)/2$, and the chi-squared approximation with the Bartlett correction is $-2\rho\log\Lambda_n \xrightarrow{d} \chi^2_f$, where $\rho = 1-\{6(k-1)(p+3)\}^{-1}(2p^2+9p+11)(\sum_{i=1}^k n_i^{-1} - n^{-1})$.

For the likelihood ratio tests (IV)–(VI), Theorem 2.2.4 gives the phase transition boundaries of the chi-squared approximations without and with the Bartlett correction.

Theorem 2.2.4 (Phase Transition Boundaries)**.** *Assume $n_i > p+1$ for $i = 1, \ldots, k$, and there exists a constant $\delta \in (0,1)$ such that $\delta < n_i/n_j < \delta^{-1}$ for any $1 \le i,j \le k$. Under H_0, for the chi-squared approximations without and with the Bartlett correction, we have the following necessary and sufficient conditions:*

(i) $\sup_{\alpha\in(0,1)} |\Pr\{-2\log\Lambda_n > \chi^2_f(\alpha)\} - \alpha| \to 0$ *if and only if* $p/n^{d_1} \to 0$;

(ii) when $p = o(n)$, $\sup_{\alpha\in(0,1)} |\Pr\{-2\rho\log\Lambda_n > \chi^2_f(\alpha)\} - \alpha| \to 0$ *if and only if* $p/n^{d_2} \to 0$,

where the values of d_1 and d_2 under the three testing problems are listed in the table below.

	(IV) Mean	*(V) Covariance*	*(VI) Joint*
(i) without correction d_1:	2/3	1/2	1/2
(ii) with correction d_2:	4/5	2/3	2/3

In Theorem 2.2.4, the boundedness of n_i/n_j suggests that the sizes of all the samples are comparable. The additional regularity condition $p = o(n)$ in (ii) specifies a local asymptotic region, which is of practical interest, and simulation studies suggest that the conclusion can hold more generally without this condition. With a fixed k, the phase transition boundaries in Theorem 2.2.4 are parallel to those in Theorem 2.2.1, and the analyses after Theorem 2.2.1 apply to Theorem 2.2.4 similarly. Particularly, examining covariances or not will yield different phase transition boundaries in the three problems. When k also increases with n, the phase transition boundaries would involve k, p, and n, as illustrated in the following proposition.

Proposition 2.2.1. *Consider $n > p + k$, $n - k \to \infty$, and $n - p \to \infty$. For Λ_n in problem (IV), under H_0, as $n \to \infty$,*

(i) $\sup_{\alpha\in(0,1)} |\Pr\{-2\log\Lambda_n > \chi^2_f(\alpha)\} - \alpha| \to 0$ *if and only if* $\sqrt{pk}(p + k)/n \to 0$;

(ii) $\sup_{\alpha\in(0,1)} |\Pr\{-2\rho\log\Lambda_n > \chi^2_f(\alpha)\} - \alpha| \to 0$ *if and only if* $\sqrt{pk}(p^2 + k^2)/n^2 \to 0$.

Proposition 2.2.1 suggests that the total number of samples k and the dimension of each observation p play symmetric roles in the phase transition boundary of problem (IV). When k is fixed, Proposition 2.2.1 is consistent with Theorem 2.2.4. To further illustrate the cases with increasing k, we consider $p = \lfloor n^\epsilon \rfloor$ and $k = \lfloor n^\eta \rfloor$, where $0 < \epsilon, \eta < 1$ and $\lfloor \cdot \rfloor$ denotes the floor of a number. Then the two phase transition boundaries in Proposition 2.2.1 become (i) $\max\{\epsilon, \eta\} + (\epsilon + \eta)/2 < 1$ and (ii) $\max\{\epsilon, \eta\} + (\epsilon + \eta)/4 < 1$, respectively. Specifically, for (i), when ϵ is close to 0, the largest value of η is around 2/3, and vice versa; when $\epsilon = \eta$, suggesting p and k are of the same order, the largest value of ϵ is 1/2. For (ii), when ϵ is close to 0, the largest value of η is around 4/5, and vice versa; when $\epsilon = \eta$, the largest value of ϵ becomes 2/3.

In addition to the phase transition boundaries above, the following Theorem 2.2.5, similarly to Theorem 2.2.2, further characterizes the accuracy of each chi-squared approximation for tests (IV)–(VI) when Wilks' theorem holds. Specifically, we consider

$p/n^{d_1} \to 0$ and $p/n^{d_2} \to 0$ for the chi-squared approximations without and with the Bartlett correction, respectively.

Theorem 2.2.5 (Asymptotic Biases). *Assume that there exists a constant $\delta \in (0,1)$ such that $\delta < n_i/n_j < \delta^{-1}$ for any $1 \le i,j \le k$, and $p \to \infty$ as $n \to \infty$. For each likelihood ratio test (IV)–(VI), let d_i, $i = 1,2$ take the corresponding values in Theorem 2.2.4. Then under H_0, for any $\alpha \in (0,1)$,*

(i) when $p/n^{d_1} \to 0$, (2.1) in Theorem 2.2.2 holds with the value of $\vartheta_1(n,p)$ below;

(ii) when $p/n^{d_2} \to 0$, (2.2) in Theorem 2.2.2 holds with the values of $\vartheta_2(n,p)$ below.

Let $D_{n,r} = \sum_{i=1}^k n_i^{-r} - n^{-r}$ and $\tilde{D}_{n,r} = \sum_{i=1}^k (n_i - 1)^{-r} - (n-k)^{-r}$.

(IV) Mean:
$$\vartheta_1(n,p) = \frac{p(k-1)(p+2+k)}{4n\sqrt{f}},$$
$$\vartheta_2(n,p) = \frac{(k-1)p(p^2+k^2-2k-4)}{24n^2\rho^2\sqrt{f}};$$

(V) Covariance:
$$\vartheta_1(n,p) = \frac{\tilde{D}_{n,1}p(2p^2+3p-1)}{24\sqrt{f}},$$
$$\vartheta_2(n,p) = \frac{p(p+1)}{24\rho^2\sqrt{f}}\Big\{(p-1)(p+2)\tilde{D}_{n,2} - 6(k-1)(1-\rho)^2\Big\};$$

(VI) Joint:
$$\vartheta_1(n,p) = \frac{D_{n,1}p\,(2p^2+9p+11)}{24\sqrt{f}},$$
$$\vartheta_2(n,p) = \frac{p(p+3)}{24\rho^2\sqrt{f}}\Big\{(p+1)(p+2)D_{n,2} - 6(k-1)(1-\rho)^2\Big\}.$$

Theorem 2.2.5 shows that for multiple-sample tests (IV)–(VI), (2.1) and (2.2) in Theorem 2.2.2 still hold. However, the values of $\vartheta_1(n,p)$ and $\vartheta_2(n,p)$ depend on the testing problems, and are different from those in Theorem 2.2.2. Similarly to Theorem 2.2.2, in each test (IV)–(VI), we also know that $\vartheta_1(n,p)$ and $\vartheta_2(n,p)$ are of the orders of $p^{1/d_1}n^{-1}$ and $p^{2/d_2}n^{-2}$, respectively. Then $\vartheta_1(n,p)\exp(-z_\alpha^2/2)/\sqrt{\pi}$ in (2.1) and $\vartheta_2(n,p)\exp(-z_\alpha^2/2)/\sqrt{\pi}$ in (2.2) are the leading terms of the biases of the chi-squared approximations without and with the Bartlett correction, respectively. We can similarly use the derived asymptotic biases to measure the approximation accuracy.

Theorem 2.2.5 focuses on the local asymptotic regime of (n, p) when Wilks' theorem holds. When p further increases such that Wilks' theorem fails, the biases of the chi-squared approximations become unignorable. The following Theorem 2.2.6 characterizes such unignorable biases of the chi-squared approximations in testing problems (IV)–(VI). Similarly to Theorem 2.2.3, we consider a general local asymptotic regime $p/n \to 0$, which includes the case when Wilks' theorem fails, i.e., $p/n^{d_1} \not\to 0$ and $p/n^{d_2} \not\to 0$ for the chi-squared approximations without and with the Bartlett correction, respectively.

Theorem 2.2.6 (Asymptotic Biases)**.** *Assume that there exists a constant $\delta \in (0,1)$ such that $\delta < n_i/n_j < \delta^{-1}$ for any $1 \leq i, j \leq k$. Moreover, assume $p \to \infty$ and $p/n_i \to 0$ as $n_i \to \infty$. For each likelihood ratio test (I)–(III), under H_0, for any $\alpha \in (0,1)$, (2.3) and (2.4) in Theorem 2.2.3 hold under three testing problems (IV)–(VI) with μ_n and σ_n listed below.*

(IV) Mean:
$$\mu_n = \frac{n}{2}\{(n-p-k-1/2)(L_{n-1,p} - L_{n-k,p}) + (k-1)L_{n-1,p} + pL_{n-1,k-1}\},$$
$$\sigma_n^2 = \frac{1}{2}(L_{n-1,p} - L_{n-k,p});$$

(V) Covariance:
$$\mu_n = \frac{1}{2}\sum_{i=1}^{k}(n_i-1)\Big\{(n-p-k-1/2)L_{n-k,p} - (n_i-p-3/2)\,L_{n_i-1,p}\Big\},$$
$$\sigma_n^2 = \frac{(n-k)^2}{2n^2}\left\{L_{n-k,p} - \sum_{i=1}^{k}\left(\frac{n_i-1}{n-k}\right)^2 L_{n_i-1,p}\right\};$$

(VI) Joint:
$$\mu_n = \frac{1}{2}\left[-kp + n\Big(n-p-\frac{3}{2}\Big)L_{n,p} - \sum_{i=1}^{k}\left\{\frac{p}{2n_i} + n_i\Big(n_i - p - \frac{3}{2}\Big)L_{n_i-1,p}\right\}\right],$$
$$\sigma_n^2 = \frac{1}{2}\left(L_{n,p} - \sum_{i=1}^{k}\frac{n_i^2}{n^2}\times L_{n_i-1,p}\right).$$

Theorem 2.2.6 shows that (2.3) and (2.4) still hold for multiple-sample tests (IV)–(VI), where the values of μ_n and σ_n^2 depend on the specific testing problem. Similarly to Theorem 2.2.3, the analysis in Remark A.3 also applies here, and we know that when $pn^{-d_1} \not\to 0$, (2.3) characterizes the unignorable biases for the chi-squared approximation, and when $pn^{-d_2} \not\to 0$, (2.4) characterizes the unignorable biases for the chi-squared approximation with the Bartlett correction. Moreover, the analysis in Remark II.1 also applies similarly to the multiple-sample tests (IV)–(VI), and thus is not repeated here.

2.2.3 Simulation Studies

We conduct simulation studies to evaluate the finite-sample performance of the theoretical results of one-sample and multiple-sample tests, respectively.

One-Sample Tests (I)–(III) Under the null hypothesis of the one-sample tests, we generate data with $\boldsymbol{\mu} = (0, \dots, 0)^\intercal$ and $\boldsymbol{\Sigma} = \mathrm{I}_p$ and use the significance level $\alpha = 0.05$.

(1) *On the phase transition boundaries.* To evaluate the phase transition boundaries in Theorem 2.2.1, we take $p = \lfloor n^\epsilon \rfloor$, where $n \in \{100, 500, 1000, 5000\}$ and $\epsilon \in \{6/24, \dots, 23/24\}$. We next plot the empirical type-I error rates (over 1000 replications) versus ϵ for each chi-squared approximation in Figure II.1.

(2) *On the asymptotic biases.* To evaluate the asymptotic biases in Theorems 2.2.2 and 2.2.3, we take $p = \lfloor n^\epsilon \rfloor$, where $n \in \{100, 500\}$ and $\epsilon \in (0, 1)$. The results of $n = 100$ and 500 (over 3000 replications) are given in Figures II.3 and II.5, respectively. In each setting, the range of ϵ is chosen such that the largest empirical type-I error is below 0.5.

To facilitate the presentation of figures and the discussions below, we define

$$\varpi_1 = \vartheta_1(n,p)\exp(-z_\alpha^2/2)/\sqrt{\pi}, \quad \varpi_3 = \bar{\Phi}\big[\big\{\chi_f^2(\alpha) + 2\mu_n\big\}/(2n\sigma_n)\big] - \alpha, \tag{2.5}$$
$$\varpi_2 = \vartheta_2(n,p)\exp(-z_\alpha^2/2)/\sqrt{\pi}, \quad \varpi_4 = \bar{\Phi}\big[\big\{\chi_f^2(\alpha) + 2\rho\mu_n\big\}/(2\rho n\sigma_n)\big] - \alpha.$$

Then $\varpi_1, \varpi_2, \varpi_3$, and ϖ_4 denote the asymptotic biases in (2.1)–(2.4), respectively. For each test in Figure II.3 and Figure II.5, we plot ϖ_1 and ϖ_2 in the subfigures in the columns (a) and (c), respectively. To better characterize each approximation bias when ϵ is beyond the corresponding phase transition boundary, we combine the results in Theorem 2.2.2 and those in Theorem 2.2.3. Specifically, in the column (b) of Figure II.3 and Figure II.5, we plot $M_c(\varpi_1, \varpi_3) \equiv \varpi_1 1\{\varpi_1 < c\} + \max\{\varpi_1, \varpi_3\}1\{\varpi_1 \geq c\}$, where $1\{\cdot\}$ denotes an indicator function, and c denotes a small positive threshold, and we choose $c = 0.002$ in the simulations. This definition of $M_c(\varpi_1, \varpi_3)$ suggests that ϖ_1 is used when the approximation bias is smaller than c, and $\max\{\varpi_1, \varpi_3\}$ is used when the approximation bias becomes larger. Similarly, we define $M_c(\varpi_2, \varpi_4) \equiv \varpi_2 1\{\varpi_2 < c\} + \max\{\varpi_2, \varpi_4\}1\{\varpi_2 \geq c\}$, and plot it in the column (d) of Figure II.3 and Figure II.5.

Remark II.2. *For each chi-squared approximation,* $\max\{\varpi_1, \varpi_3\}$ *already characterizes the bias well most of the time. We use* $M_c(\varpi_1, \varpi_3)$ *instead of* $\max\{\varpi_1, \varpi_3\}$ *because* ϖ_3 *can mistakenly indicate a large bias under small* ϵ*, especially when* n *is small. Compared to* $\max\{\varpi_1, \varpi_3\}$*,* $M_c(\varpi_1, \varpi_3)$ *does not use* ϖ_3 *when* ϖ_1 *indicates that the bias is still small. As long as* c *is sufficiently small but not too close to zero,* $M_c(\varpi_1, \varpi_3)$ *will not take the wrong value given by* ϖ_3*, and thus gives a good evaluation of the approximation bias under a wide range of* ϵ *values. Despite the difference between* $M_c(\varpi_1, \varpi_3)$ *and* $\max\{\varpi_1, \varpi_3\}$*, we note that* $M_c(\varpi_1, \varpi_3)$ *is equal to* $\max\{\varpi_1, \varpi_3\}$ *under most cases. For instance, in all our simulations with* $n = 500$ *and* $c = 0.002$*,* $M_c(\varpi_1, \varpi_3) = \max\{\varpi_1, \varpi_3\}$*. When the Bartlett correction is used, we know that similar analysis applies to* $\max\{\varpi_2, \varpi_4\}$ *and* $M_c(\varpi_2, \varpi_4)$*.*

Multiple-Sample Tests (IV)–(VI) Consider $k = 3$, $n_1 = n_2 = n_3$, and $n = n_1 + n_2 + n_3$. Under the null hypothesis of each multiple-sample test (IV)–(VI), we set $\boldsymbol{\mu}_i = (0, \ldots, 0)^\intercal$, and $\boldsymbol{\Sigma}_i = \mathrm{I}_p$ for $i = 1, 2, 3$.

(1) *On the phase transition boundaries.* Let $p = \lfloor n^{\epsilon} \rfloor$, where $n = n_1 + n_2 + n_3$ and $n_i \in \{100, 500, 1000, 5000\}$ for $i = 1, 2, 3$. We then plot the empirical type-I error rates (over 1000 replications) versus ϵ for each chi-squared approximation in Figure II.2.

(2) *On the asymptotic biases.* To evaluate the asymptotic biases in Theorems 2.2.5 and 2.2.6, we take $p = \lfloor n^{\epsilon} \rfloor$, where $n = n_1 + n_2 + n_3$, $n_i \in \{100, 500\}$ for $i = 1, 2, 3$, and $\epsilon \in (0, 1)$. The results of $n_i = 100$ and 500 (over 3000 replications) are given in Figure II.6 and Figure II.7, respectively. Similarly to Figure II.3 and Figure II.5, in each row of Figure II.6 and Figure II.7, the lines with dot markers in the four columns (a)–(d) give ϖ_1, $M_c(\varpi_1, \varpi_3)$, ϖ_2, and $M_c(\varpi_2, \varpi_4)$, respectively.

We next analyze the simulation results. First, as shown in Figures II.1 and II.2, the theoretical phase transition boundary, denoted by a vertical line, is observed to be consistent with where each chi-squared approximation starts to fail. For instance, the two plots in the first row of Figure II.1 show that for test (I), the type-I error rates of the chi-squared approximations without and with the Bartlett correction begin to inflate when ϵ is around 2/3 and 4/5, respectively. These are consistent with $d_1 = 2/3$ and $d_2 = 4/5$ for test (I) in Theorem 2.2.1. Similarly for other tests, we can see that the numerical results are also consistent with the corresponding conclusions in Theorems 2.2.1 and 2.2.4.

Second, Figures II.3–II.7 show that the derived theoretical asymptotic biases provide good evaluations of the corresponding chi-squared approximation biases. From the subfigures in the column (a) of Figures II.3–II.7, we can see that as ϵ increases, the empirical type-I error inflates, and ϖ_1 also increases accordingly. At the ϵ values where the type-I error begins to inflate, the difference between the empirical type-I error and ϖ_1 is close to 0.05, as shown by the circle line, which suggests that ϖ_1 approximates the chi-squared approximation bias $\Pr\{-2\log\Lambda_n > \chi^2_f(\alpha)\} - \alpha$ well in this regime. When ϵ further increases beyond the corresponding phase transition boundary, the asymptotic bias ϖ_1 keeps increasing, and its large value indicates the failure of the chi-squared approximation, even though now ϖ_1 underestimates the approximation bias in this regime. To better characterize the approximation bias when ϵ is beyond the phase transition boundary, we combine ϖ_1 and ϖ_3 by plotting $M_c(\varpi_1, \varpi_3)$ in the column (b) of Figures II.3–II.7. The results suggest that utilizing the two asymptotic biases in (2.1) and in (2.3) together can give a good evaluation of the approximation bias under a wide range of ϵ values, either below or above the phase transition boundary. Moreover, in each subfigure in the column (b), we also highlight the location with x-axis ϵ^* where $M_c(\varpi_1, \varpi_3)$ starts to be larger than ϖ_1 (the plus sign). When $\epsilon < \epsilon^*$, $M_c(\varpi_1, \varpi_3) = \varpi_1$, indicating that ϖ_1 approximates the bias better than ϖ_3 does in this regime, while ϖ_3 performs better than ϖ_1 when $\epsilon \geq \epsilon^*$. Similarly, for the chi-squared approximation with the Bartlett correction, similar conclusions can be obtained by the results in the columns (c) and (d) of Figures II.3–II.7.

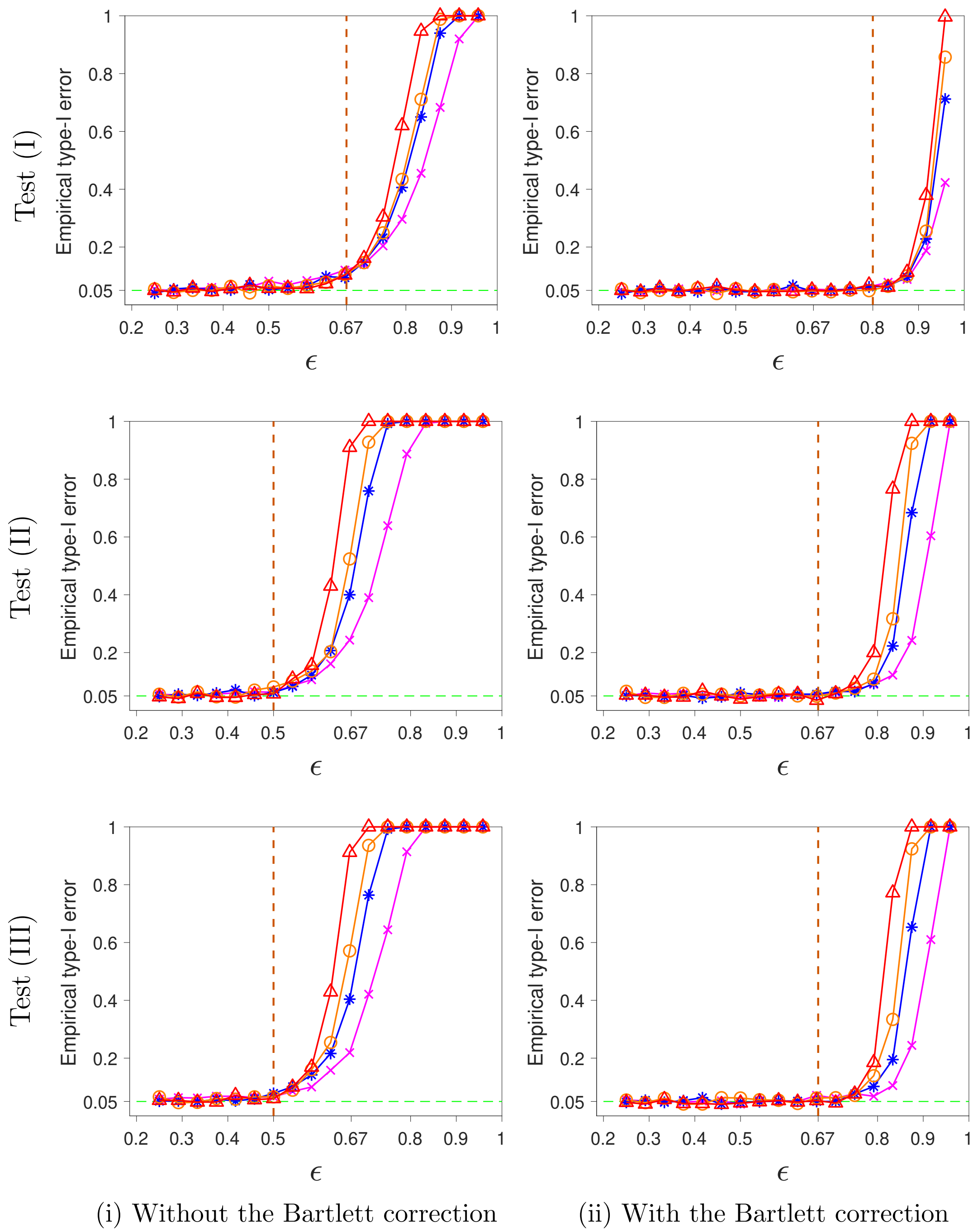


Figure II.1: Illustration of phase transitions of one-sample tests (I)–(III): Empirical type I errors versus ϵ when $p = \lfloor n^{\epsilon} \rfloor$ for $n = 100$ (cross), 500 (asterisk), 1000 (square), and 5000 (triangle); theoretical phase transition boundary (vertical dashed line).

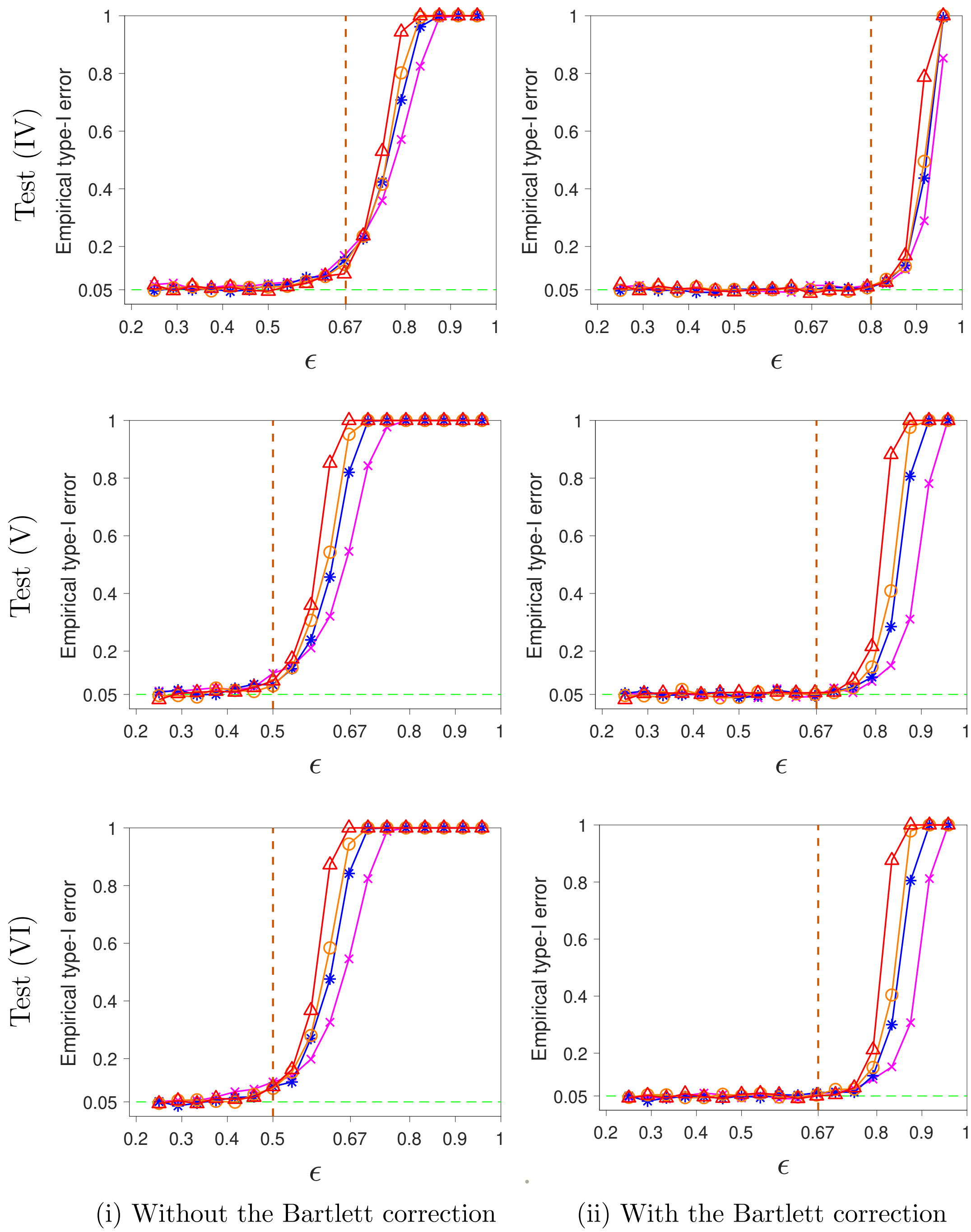
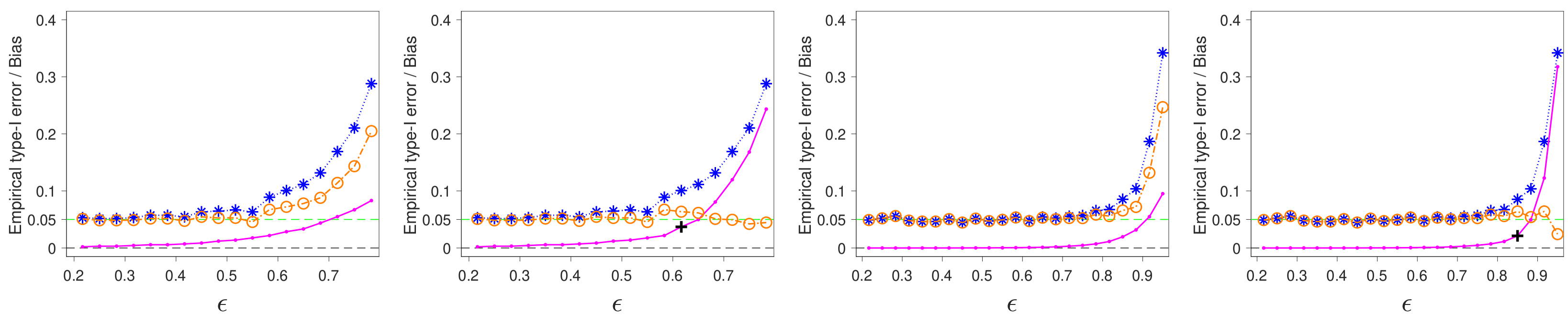


Figure II.2: Illustration of phase transitions of multiple-sample tests (IV)–(VI): Empirical type I errors versus ϵ when $p = \lfloor n^{\epsilon} \rfloor$ for $n = 100$ (cross), 500 (asterisk), 1000 (square), and 5000 (triangle); theoretical phase transition boundary (vertical dashed line).

Test (I)



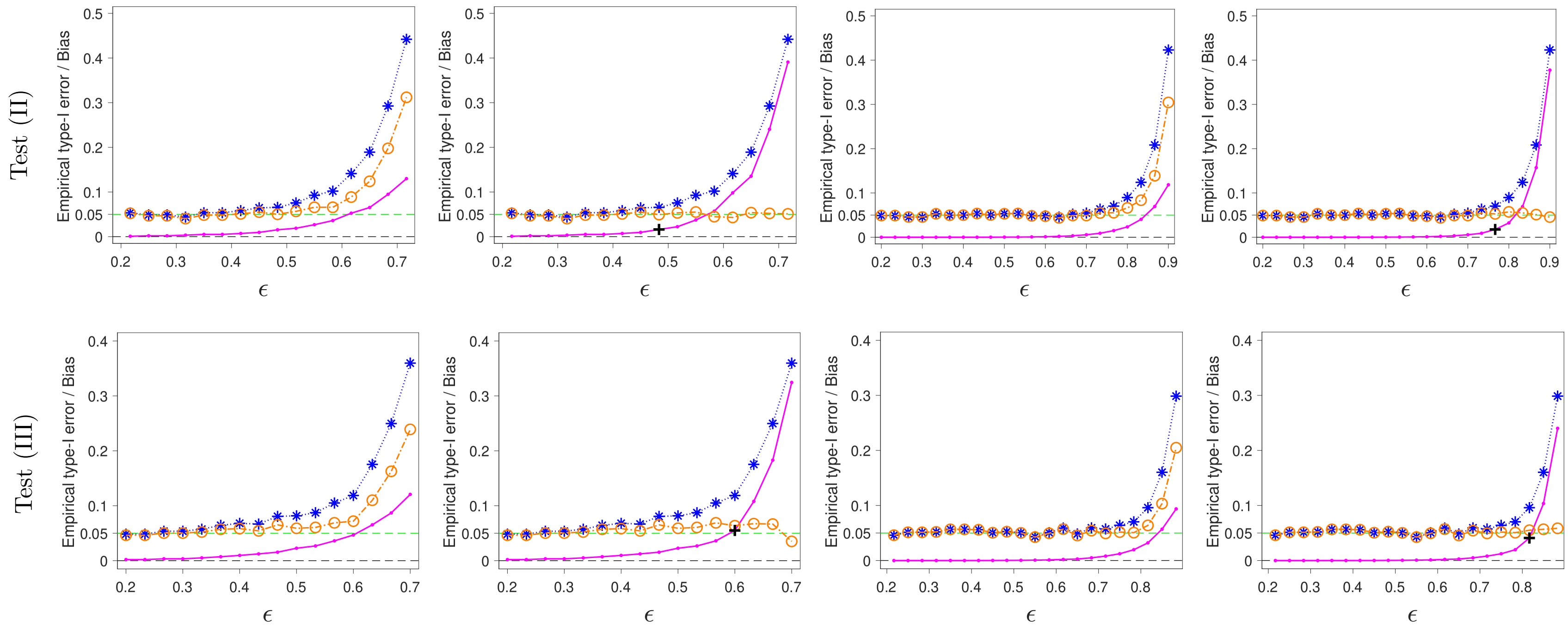
(a) Without the Bartlett correction

(b) Without the Bartlett correction

(c) With the Bartlett correction

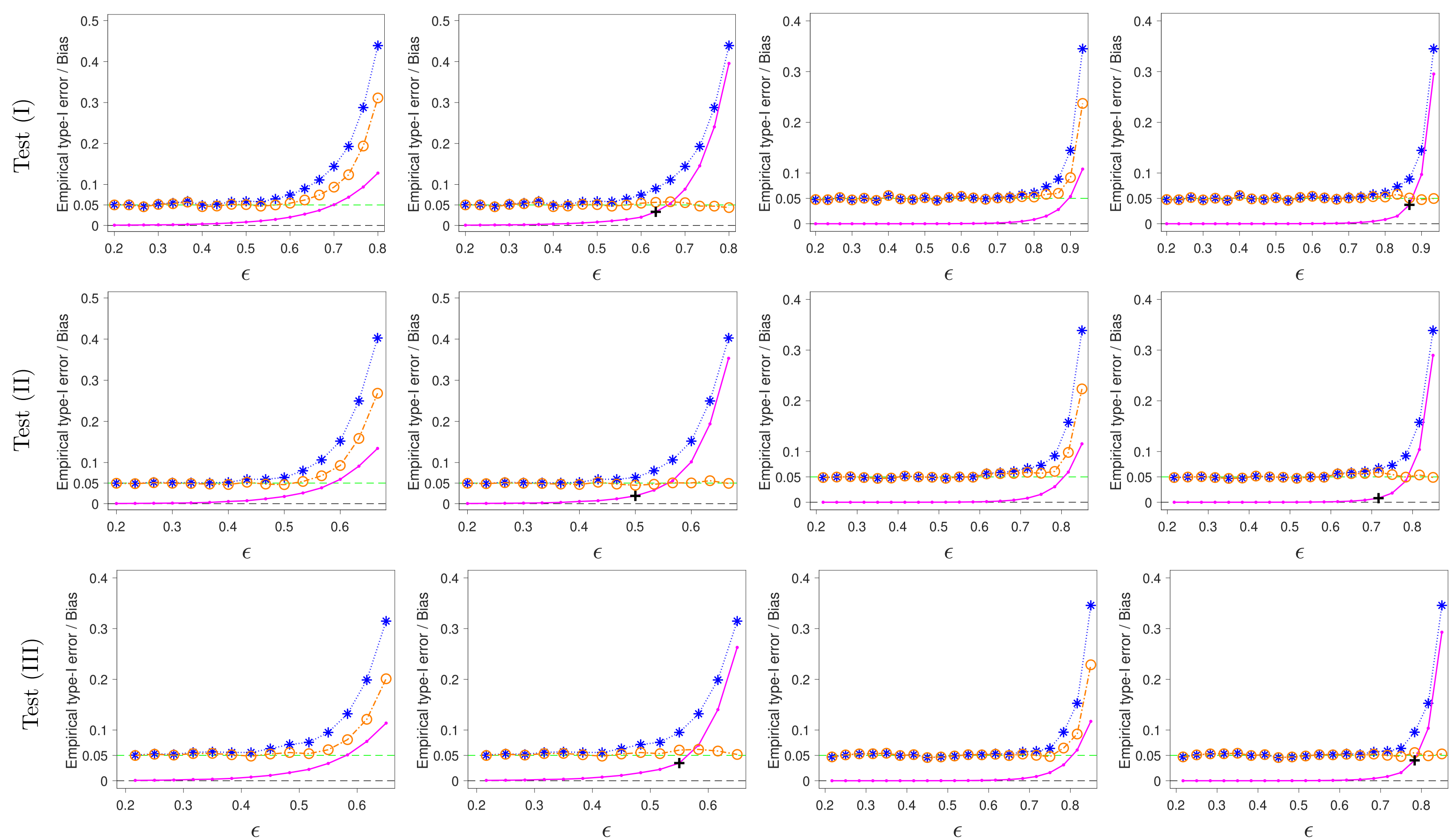
(d) With the Bartlett correction

Figure II.3: Empirical type-I errors and the theoretical asymptotic biases of the one-sample mean test (I) versus ϵ when $p = \lfloor n^{\epsilon} \rfloor$ and $n = 100$. (a): empirical type-I errors of the test without the Bartlett correction (asterisk); ϖ_1, i.e., the asymptotic bias in Eq. (2.1) (dot); the difference between the empirical type-I errors and ϖ_1 (circle). (b): empirical type-I errors of the test without the Bartlett correction (asterisk); $M_c(\varpi_1, \varpi_3)$ with $c = 0.002$ (dot); the location with x-axis ϵ^* satisfying $M_c(\varpi_1, \varpi_3) = \varpi_1$ when $\epsilon < \epsilon^*$ and $M_c(\varpi_1, \varpi_3) > \varpi_1$ when $\epsilon \geq \epsilon^*$ (plus sign); the difference between the empirical type-I error and $M_c(\varpi_1, \varpi_3)$ (circle). (c): empirical type-I errors of the test with the Bartlett correction (asterisk); ϖ_2, i.e., the asymptotic bias in Eq. (2.2) (dot); the difference between the empirical type-I error and ϖ_2 (circle). (d): empirical type-I errors of the test with the Bartlett correction (asterisk); $M_c(\varpi_2, \varpi_4)$ with $c = 0.002$ (dot); the location with x-axis ϵ^* satisfying $M_c(\varpi_2, \varpi_4) = \varpi_2$ when $\epsilon < \epsilon^*$ and $M_c(\varpi_2, \varpi_4) > \varpi_2$ when $\epsilon \geq \epsilon^*$ (plus sign); the difference between the empirical type-I errors and $M_c(\varpi_2, \varpi_4)$ (circle).



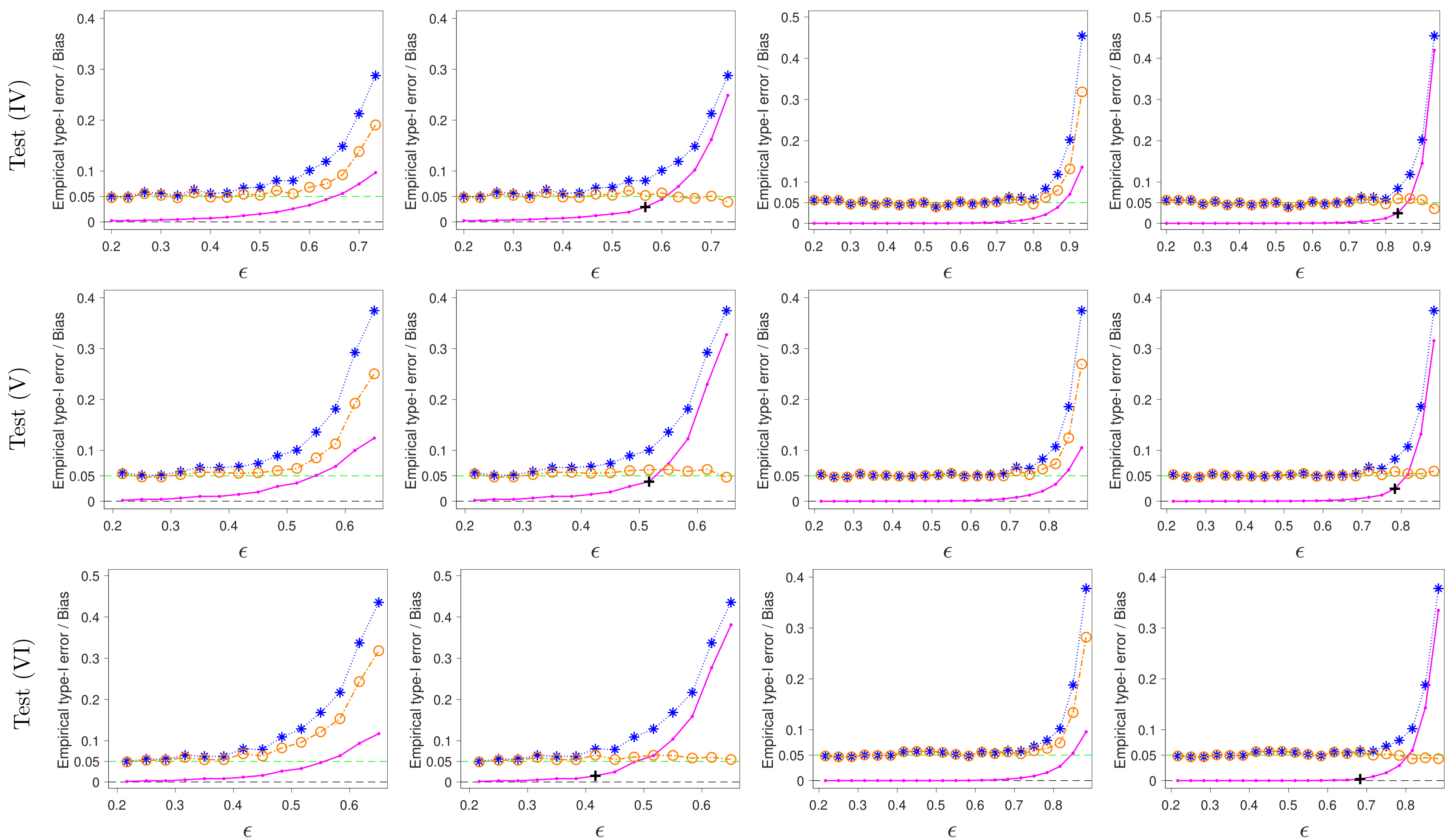
(a) Without the Bartlett correction (b) Without the Bartlett correction (c) With the Bartlett correction (d) With the Bartlett correction

Figure II.4: Empirical type-I errors and the theoretical asymptotic biases of the one-sample tests (II)–(III) versus ϵ when $p = \lfloor n^{\epsilon} \rfloor$ and $n = 100$. (Columns (a)–(d): see the caption description in Figure II.3.)



(a) Without the Bartlett correction (b) Without the Bartlett correction (c) With the Bartlett correction (d) With the Bartlett correction

Figure II.5: Empirical type-I errors and the theoretical asymptotic biases of the one-sample tests (I)–(III) versus ϵ when $p = \lfloor n^{\epsilon} \rfloor$ and $n = 500$. (Columns (a)–(d): see the caption description in Figure II.3.)



(a) Without the Bartlett correction (b) Without the Bartlett correction (c) With the Bartlett correction (d) With the Bartlett correction

Figure II.6: Empirical type-I errors and the theoretical asymptotic biases of the multiple-sample tests (IV)–(VI) versus ϵ when $p = \lfloor n^{\epsilon} \rfloor$ and $n = 100$. (Columns (a)–(d): see the caption description in Figure II.3.)

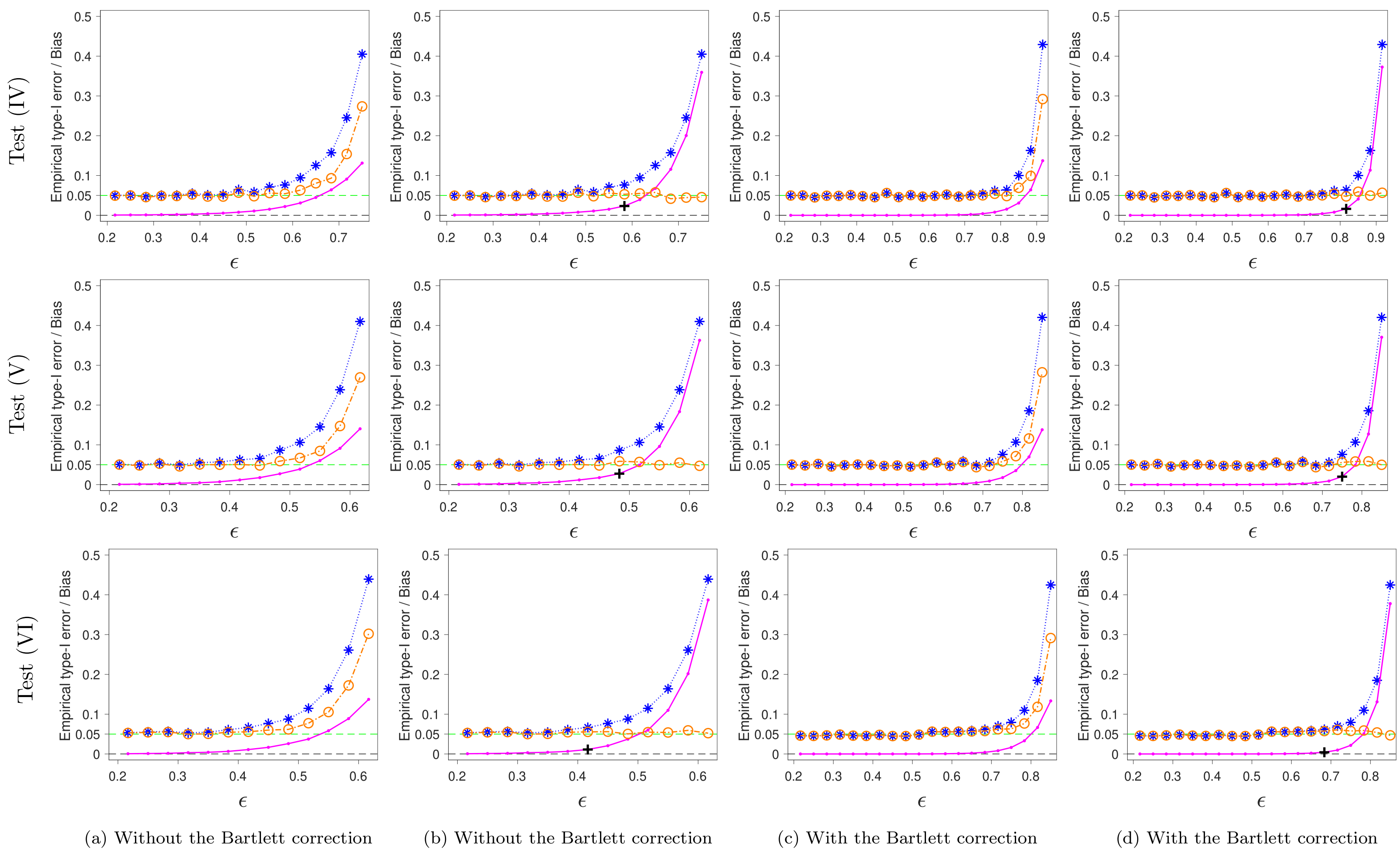


Figure II.7: Empirical type-I errors and the theoretical asymptotic biases of the multiple-sample tests (IV)–(VI) versus ϵ when $p = \lfloor n^{\epsilon} \rfloor$ and $n = 500$. (Columns (a)–(d): see the caption description in Figure II.3.)

2.3 Results for Exploratory Factor Analysis

Exploratory factor analysis serves as a popular statistical tool to gain insights into latent structures underlying the observed data (Gorsuch, 1988; Fabrigar and Wegener, 2011; Bartholomew et al., 2011). It is widely used in many application areas such as psychological and social sciences (Fabrigar et al., 1999; Preacher and MacCallum, 2002; Thompson, 2004; Finch and Finch, 2016). In factor analysis, the relationship among observed variables in data are explained by a smaller number of unobserved underlying variables, called common factors. To understand the underlying scientific patterns, one fundamental problem in factor analysis is to decide the minimum number of latent common factors that is needed to describe the statistical dependencies in data.

In order to determine the number of factors in exploratory factor analysis, a wide variety of procedures have been proposed; see reviews and discussions in Costello and Osborne (2005), Barendse et al. (2015) and Luo et al. (2019). For instance, one broad class of criteria are based on the eigenvalues of the sample correlation matrix of the observed data. Examples include the Kaiser criterion (Kaiser, 1960), the scree test (Cattell, 1966), the parallel analysis method (Horn, 1965; Keeling, 2000; Dobriban, 2020), and testing linear trend of eigenvalues (Bentler and Yuan, 1998) among many others. Another class of methods propose various goodness-of-fit indexes to select the number of factors, such as AIC (Akaike, 1987), BIC (Schwarz, 1978), the reliability coefficient (Tucker and Lewis, 1973), and the root mean square error of approximation (Steiger, 2016). Moreover, the likelihood ratio test provides another popularly used approach in practice (Bartlett, 1950; Anderson, 2003).

Among the various criteria to determine the number of factors, the likelihood ratio test plays a unique role, as it is based on a formal hypothesis testing procedure with a clear statistical rationale and also has a solid theoretical foundation with guaranteed statistical properties. In particular, the likelihood ratio test examines how a factor analysis model fits the data using a hypothesis testing framework based on the likelihood theory. The classical statistical theory shows that under the null hypothesis, the likelihood ratio test statistic (after proper scaling) asymptotically converges to a chi-squared distribution, with the degrees of freedom equal to the difference in the number of free parameters between the null and alternative hypothesis models (see, e.g., Anderson, 2003, Section 14.3.2).

In the modern big data era, it is of emerging interest to analyze high-dimensional data (Finch and Finch, 2016; Harlow and Oswald, 2016; Chen et al., 2019b), where throughout this section we refer to the dimension of the observed response variables as the dimension of data. Classical asymptotic theory, despite its importance, often replies on the assumption that the data dimension is fixed as the sample size increases. Such an assumption often fails in high-dimensional data analysis with large data dimension, and therefore the corresponding asymptotic theory is no longer directly applicable to modern high-dimensional applications. In fact, it has been found in the recent statistical literature that the chi-squared approximations for the likelihood ratio test statistics can become inaccurate as the dimension of data increases with the sample size (e.g. Bai et al., 2009; Jiang and Yang, 2013; He et al., 2021a). In factor analysis, although considerable high-dimensional statistical analysis results have been recently developed (Bai and Ng, 2002; Bai and Li, 2012; Sundberg and Feldmann, 2016; Ait-Sahalia and Xiu, 2017; Chen and Li, 2020), less attention has been paid to the statistical properties of the popular likelihood ratio test under high dimensions. Particularly, it remains an open problem when the conventional chi-squared approximation of the likelihood ratio test starts to fail as the data dimension grows. In other words, for a dataset with sample size N, how large the data dimension p can be to still ensure the validity of the chi-squared approximation of the likelihood ratio test?

To better understand this issue, the remaining part of this section investigates the influence of the data dimensionality on the likelihood ratio test in high-dimensional exploratory factor analysis. Particularly, Section 2.3.1 gives a brief review of the exploratory factor analysis and the likelihood ratio test, and Section 2.3.2 presents our theoretical and numerical results on the performance of the chi-squared approximation under high dimensions. The technical proofs are deferred to the appendix.

2.3.1 Likelihood Ratio Test in Exploratory Factor Analysis

In this section, we briefly review the likelihood ratio test in exploratory factor analysis (see, e.g., Anderson, 2003, Section 14). Suppose X_i, $i = 1, \ldots, N$ are independent and identically distributed p-dimensional random vectors. The exploratory factor analysis considers the following common-factor model

$$X_i = \mu + \Lambda F_i + U_i, \tag{2.6}$$

where μ a the p-dimensional mean parameter vector, Λ is a $p \times k_0$ loading matrix with $\mathrm{rank}(\Lambda) = k_0 < p$, F_i is a k_0-dimensional random vector containing the common factors, and U_i is a p-dimensional error vector. It is well known that the factor model (2.6) is not identifiable without additional constraints, and there are many ways to impose identifiablity restrictions (Anderson, 2003; Bai and Li, 2012). In this section, we focus on the following identification conditions which have been popularly used in exploratory factor analysis. In particular, we assume that F_i and U_i are independent latent random vectors with $\mathrm{E}(F_i) = \mathbf{0}_{k_0}$, $\mathrm{cov}(F_i) = \mathrm{I}_{k_0}$, $\mathrm{E}(U_i) = \mathbf{0}_p$, and $\mathrm{cov}(U_i) = \Psi$, where $\mathbf{0}_{k_0}$ denotes a k_0-dimensional all-zero vector, I_{k_0} represents a $k_0 \times k_0$ identity matrix, and Ψ is a $p \times p$ diagonal matrix with $\mathrm{rank}(\Psi) = p$. It follows that the population covariance matrix $\Sigma = \mathrm{cov}(X_i)$ can be expressed as

$$\Sigma = \Lambda\Lambda^\top + \Psi. \tag{2.7}$$

Typically, the true number of common factors k_0 is unknown. In exploratory factor analysis, to determine the number of factors in model (2.6), various procedures have been developed. Among them, the likelihood ratio test plays a unique role due to its solid theoretical foundation and nice statistical properties. The common practice utilizes the model's likelihood function assuming both F_i and U_i to be normally distributed. In such case, X_i follows a multivariate normal distribution with mean vector $\mathbf{0}_p$ and covariance matrix Σ as in (2.7), and we write $X_i \sim \mathcal{N}(\mathbf{0}_p, \Sigma)$. Then, the likelihood ratio test is used to sequentially test the factor analysis model with a specified number of factors against the saturated model (e.g., Hayashi et al., 2007).

Specifically, for each $k = 0, 1, \dots, p$, we consider the following null and alternative hypotheses: $H_{0,k} : \Sigma = \Lambda\Lambda^\top + \Psi$ with (at most) k factors, versus $H_{A,k} : \Sigma$ is any positive definite matrix. In practice without a priori knowledge, a typical procedure examines the above hypotheses in a forward stepwise manner. Specifically, we first consider $k = 0$ and examine $H_{0,0} : k_0 = 0$ versus $H_{A,0}$ using the likelihood ratio test, that is, testing whether there is any factor in model (2.6). If $H_{0,0}$ is rejected, we then consider $k = 1$, that is, a 1-factor model in the null hypothesis $H_{0,1}$. If $H_{0,1}$ is rejected, we proceed with $k = 2$, and test a 2-factor model for $H_{0,2}$. This testing procedure continues until we fail to reject $H_{0,\hat{k}}$ for some $\hat{k}$. Then $\hat{k}$ is taken as an estimate of the true number of factors based on the likelihood ratio test.

We next introduce the details on the abovementioned likelihood ratio test. For $k = 0$, $H_{0,0}$ examines the existence of any significant factors, which is an important problem in psychology applications (e.g., Mukherjee, 1970). This test can be written as

$$H_0 : \Sigma = \Psi \quad \text{versus} \quad H_A : \Sigma \neq \Psi,$$

that is, testing whether Σ is a diagonal matrix. Statistically, this is also equivalent to the following hypothesis test

$$H_0 : R = \mathrm{I}_p \quad \text{versus} \quad H_A : R \neq \mathrm{I}_p,$$

where R denotes the population correlation matrix of the response variables $\{X_i, i = 1, \ldots, N\}$. Under the normality assumption of X, $H_{0,0}$ then tests for the complete independence between p dimensions of X. The likelihood ratio test statistic for $H_{0,0}$ with the chi-squared limit is $T_0 = -(N-1)\log(|\hat{R}_N|)$, where $\hat{R}_N$ denotes the sample correlation matrix of the observations $\{X_i, i = 1, \ldots, N\}$, and $|\hat{R}_N|$ denotes the determinant of $\hat{R}_N$; see, e.g., Bartlett (1950). When the dimension p is fixed and the sample size $N \to \infty$, under $H_{0,0}$,

$$T_0 \xrightarrow{D} \chi^2_{f_0}, \quad \text{with } f_0 = p(p-1)/2, \tag{2.8}$$

where $\xrightarrow{D}$ represents the convergence in distribution, and $\chi^2_{f_0}$ represents a random variable following the chi-squared distribution with degrees of freedom f_0. To improve the finite-sample performance, researchers have proposed using the Bartlett correction for the likelihood ratio test (Bartlett, 1950). The corrected test statistic is $\rho_0 T_0$ with the Bartlett correction term $\rho_0 = 1 - (2p+5)/\{6(N-1)\}$, and under $H_{0,0}$ with fixed p and $N \to \infty$, we still have the chi-squared approximation:

$$\rho_0 \times T_0 \xrightarrow{D} \chi^2_{f_0}, \tag{2.9}$$

while it improves the convergence rate of the chi-squared approximation (3) from $O(N^{-1})$ to $O(N^{-2})$.

For $k \geq 1$, $H_{0,k}$ examines whether the k-factor model fits the observed data. Under the k-factor model, let $\hat{\Lambda}_k$ and $\hat{\Psi}_k$ denote the maximum likelihood estimators of Λ and Ψ, respectively, and define $\hat{\Sigma}_k = \hat{\Lambda}_k \hat{\Lambda}_k^\top + \hat{\Psi}_k$. Then to test $H_{0,k}$, the likelihood ratio test statistic can be written as

$$T_k = -(N-1)\log(|\hat{\Sigma}| \times |\hat{\Sigma}_k|^{-1}) + (N-1)\{\mathrm{tr}(\hat{\Sigma}\hat{\Sigma}_k^{-1}) - p\}, \tag{2.10}$$

where $\hat{\Sigma}$ is the unbiased sample covariance matrix of the observations $\{X_i, i = 1, \cdots, N\}$, and $\mathrm{tr}(A)$ denotes the trace of a matrix A; see, e.g., Lawley and Maxwell (1962). Under the null hypothesis with $k_0 = k$, p fixed and $N \to \infty$, we have the following chi-squared approximation:

$$T_k \xrightarrow{D} \chi^2_{f_k}, \quad \text{where} \;\; f_k = \{(p-k)^2 - p - k\}/2. \tag{2.11}$$

Moreover, applying the Bartlett correction for this test, we have

$$\rho_k \times T_k \xrightarrow{D} \chi^2_{f_k}, \quad \text{where } \rho_k = 1 - \frac{2p + 5 + 4k}{6(N-1)}. \tag{2.12}$$

Despite the usefulness of the above chi-squared approximations, classical large sample theory assumes that the data dimension p is fixed, and therefore many conclusions are not directly applicable to high-dimensional data when p increases with the sample size N. As analyzing high-dimensional data is of emerging interest in modern data science, it imposes new challenges to understanding the statistical performance of the likelihood ratio test in the exploratory factor analysis, which will be investigated in the next section.

2.3.2 Phase Transition Boundary

In high-dimensional exploratory factor analysis, it is important to understand the limiting behavior of the likelihood ratio test, as applying an inaccurate limiting distribution would lead to misleading scientific conclusions. This section focuses on the limiting distribution of the likelihood ratio test under the null hypothesis, and investigates the influence of the data dimension p and the sample size N on the chi-squared approximation.

Recent statistical literature has shown that the chi-squared approximation for the likelihood ratio test can become inaccurate in various testing problems (Bai et al., 2009; Jiang and Yang, 2013; He et al., 2021a), while this inaccuracy issue is still less studied in the exploratory factor analysis. To demonstrate that similar phenomena exist for the exploratory factor analysis, we first present a numerical example, before showing our theoretical results.

Numerical Example 1. *Consider $H_{0,0}$ in Section 2.3.1 with $N = 1000$ and $p \in \{20, 100, 300, 500\}$. Under each combination of (N, p), we generate $X_i, i = 1, \ldots, N$ from $\mathcal{N}(\mathbf{0}_p, \mathrm{I}_p)$ independently, and then compute the likelihood ratio test statistics T_0 in (2.8) and its Bartlett corrected version $\rho_0 T_0$ in (2.9). We repeat the procedure 5000 times, and present the histograms of T_0 and $\rho_0 T_0$ in the first and second rows, respectively, of Figure II.8. For comparison, in each histogram, we add the theoretical density curve of the limiting distribution $\chi^2_{f_0}$ in (2.8) and (2.9) (the red curves in Figure II.8).*

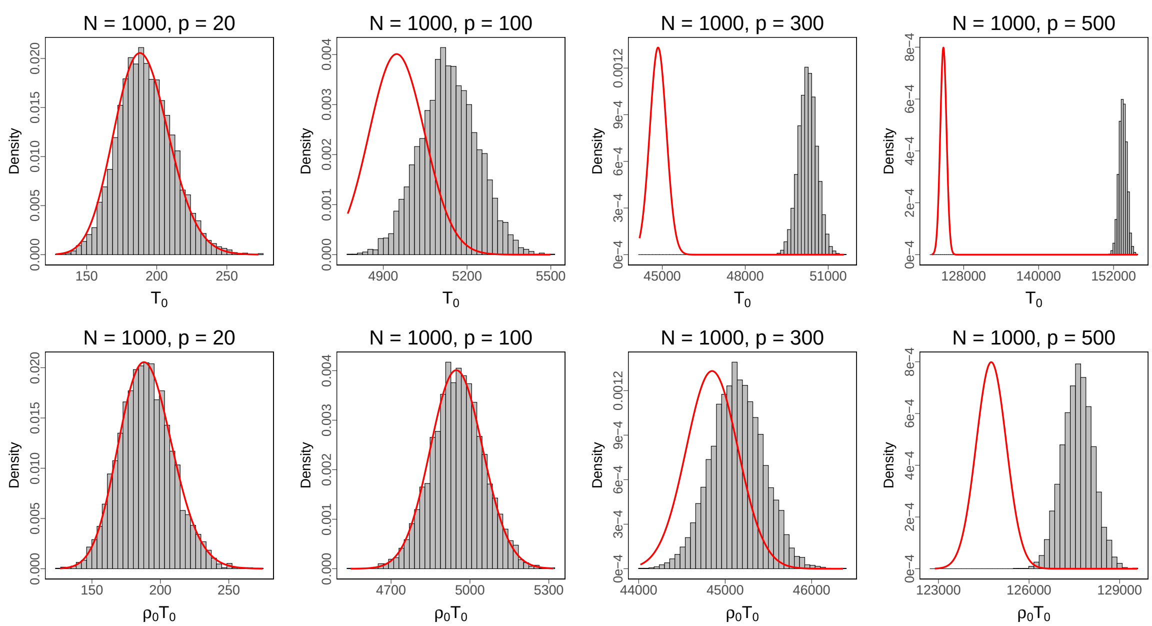


Figure II.8: Histograms of T_0 and $\rho_0 T_0$ and the density curves of $\chi^2_{f_0}$ (the red curve) in the numerical example 1.

From the two figures in the first column of Figure II.8, we can see that when p is small ($p = 20$) compared to N, the density curve of $\chi^2_{f_0}$ approximates the histograms of T_0 and $\rho_0 T_0$ well. This is consistent with the classical large sample theory in (2.8) and

(2.9). However, as p increases from 20 to 500, the density curve of $\chi^2_{f_0}$ moves farther away from the sample histograms of T_0 and $\rho_0 T_0$, indicating the failure of the chi-squared approximation as p increases. It is also interesting to note that the likelihood ratio test statistics without and with the Bartlett correction behave differently as p increases, despite their similarity when p is small. For instance, when $p = 100$, $\chi^2_{f_0}$ already fails to approximate the distribution of T_0, but it can still well approximate that of the corrected statistic $\rho_0 T_0$. Nevertheless, when $p = 300$ and 500, $\chi^2_{f_0}$ fails to approximate the distributions of both T_0 and $\rho_0 T_0$, while the approximation biases differ. These numerical observations bring the following question in practice: how large the dimension p with respect to the sample size N can be so that we can still apply the classic chi-squared approximation for the likelihood ratio test?

To provide a statistical insight into this important practical issue, we derive the necessary and sufficient condition to ensure the validity of the chi-squared approximation for the likelihood ratio test, as p increases with N. Particularly, we first consider $H_{0,0} : k_0 = 0$ in Section 2.3.1 and provide the following Theorem 2.3.1.

Theorem 2.3.1. *Suppose $N \geq p + 5$. Let $\chi^2_{f_0}(\alpha)$ denote the upper-level α-quantile of the $\chi^2_{f_0}$ distribution. Under $H_{0,0} : k_0 = 0$, as $N \to \infty$,*

(i) $\sup_{\alpha \in (0,1)} |\Pr\{T_0 > \chi^2_{f_0}(\alpha)\} - \alpha| \to 0$, *if and only if* $\lim_{n\to\infty} p/N^{1/2} = 0$*;*

(ii) $\sup_{\alpha \in (0,1)} |\Pr\{\rho_0 \times T_0 > \chi^2_{f_0}(\alpha)\} - \alpha| \to 0$, *if and only if* $\lim_{n\to\infty} p/N^{2/3} = 0$.

In Theorem 2.3.1, $N \geq p + 5$ is required for the technical proof. This condition is mild as $N \geq p + 1$ is required for the existence of the likelihood ratio test statistic with probability one (Jiang and Yang, 2013). Theorem 2.3.1 (i) suggests that the chi-squared approximation for T_0 in (2.8) starts to fail when the dimension p approaches $N^{1/2}$, and (ii) shows that the chi-squared approximation for $\rho_0 T_0$ in (2.9) starts to fail when p approaches $N^{2/3}$. To further demonstrate the validity of Theorem 2.3.1, we conduct a simulation study as follows.

Numerical Example 2. *We take $p = \lfloor N^{\varepsilon} \rfloor$, where $N \in \{100, 500, 1000, 2000\}$ and $\varepsilon \in \{3/24, 4/24, \ldots, 23/24\}$. For each combination of (N, p), we generate X_i from $\mathcal{N}(\mathbf{0}_p, \mathrm{I}_p)$ for $i = 1, \ldots, N$ independently, and conduct the likelihood ratio test with two chi-squared approximations in (2.8) and (2.9), respectively. We repeat the procedure 1000 times to estimate the type I error rates with significance level 0.05, and then plot estimated type I error rates versus ε in Figure II.9. The left figure in Figure II.9 presents the results of the chi-squared approximation for T_0 in (2.8), where the estimated type I error begins to inflate when ε approaches $1/2$. In addition, the right figure in Figure II.9 presents the results of the chi-squared approximation for $\rho_0 T_0$ in (2.9), where the estimated type I error begins to inflate when ε approaches $2/3$. The two theoretical boundaries on ε in Theorem 2.3.1 are denoted by two vertical dashed lines in Figure II.9. For each approximation, the theoretical and empirical values of ε where the approximation begins to fail are consistent.*

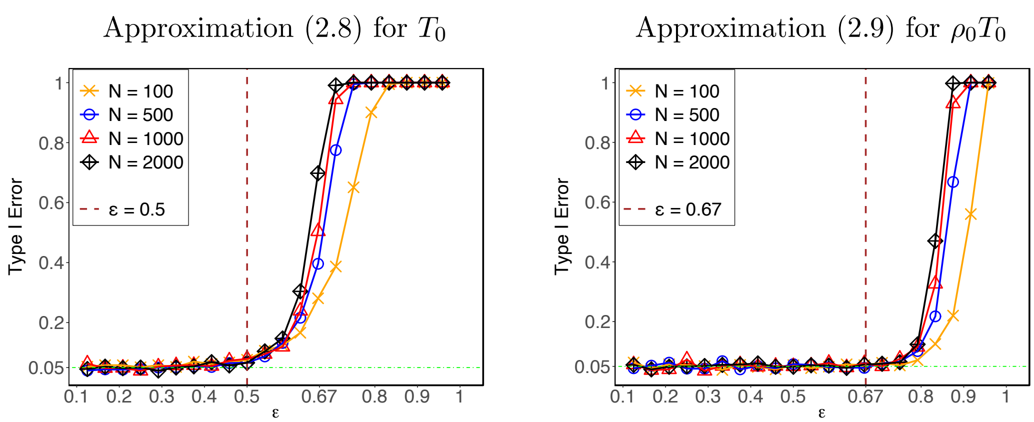


Figure II.9: Empirical type I errors versus ε when $k_0 = 0$ in the numerical example 2.

We next investigate the sequential test for $H_{0,k}$ when $k \geq 1$. Under $H_{0,k}$, assume the true factor number is k, and $\Lambda_k \Lambda_k^{\top}$ and Ψ_k are the true values such that (2.7) holds with $\Lambda\Lambda^{\top} = \Lambda_k \Lambda_k^{\top}$ and $\Psi = \Psi_k$, where Λ_k is a matrix of size $p \times k$, and Ψ_k is a diagonal matrix. In classical multivariate analysis with fixed dimension and certain regularity conditions, it can be shown that $\hat{\Lambda}_k\hat{\Lambda}_k^\top \xrightarrow{P} \Lambda_k\Lambda_k^\top$ and $\hat{\Psi}_k \xrightarrow{P} \Psi_k$, where $\xrightarrow{P}$ represents the convergence in probability; see, e.g., Theorem 14.3.1 in Anderson (2003). To facilitate the following theoretical analysis, we consider a simplified version of the test by assuming $\Lambda_k\Lambda_k^\top$ and Ψ_k are given, and define $\Sigma_k = \Lambda_k\Lambda_k^\top + \Psi_k$. Then we consider testing $H'_{0,k} : \Sigma = \Sigma_k$, and the likelihood ratio test statistic can be expressed as

$$T' = -(N-1)\log(|\hat{\Sigma}| \times |\Sigma_k|^{-1}) + (N-1)\{\mathrm{tr}(\hat{\Sigma}\Sigma_k^{-1}) - p\};$$

see Section 8.4 of Muirhead (2009). The test statistic T' and T_k in (2.10) are the same except that T' is based on the true value $\Sigma_k = \Lambda_k\Lambda_k^\top + \Psi_k$, while T_k is based on $\hat{\Sigma}_k = \hat{\Lambda}_k\hat{\Lambda}_k^\top + \hat{\Psi}_k$, with $\hat{\Lambda}_k\hat{\Lambda}_k^\top$ and $\hat{\Psi}_k$ being the maximum likelihood estimators of $\Lambda_k\Lambda_k^\top$ and Ψ_k, respectively, under the k-factor model. Under the classical setting with p fixed, the chi-squared approximation of T' is $T' \xrightarrow{D} \chi^2_{f'}$, where $f' = p(p+1)/2$, and by the Bartlett correction with $\rho' = 1 - \{6(N-1)(p+1)\}^{-1}(2p^2+3p-1)$, we have $\rho' T' \xrightarrow{D} \chi^2_{f'}$. For this simplified testing problem $H'_{0,k}$, the test statistic T' and its limit do not depend on the number of factors k, as the true $\Lambda_k\Lambda_k^\top$ and Ψ_k are assumed to be given.

Considering $H'_{0,k}$ and the statistic T', we next provide the necessary and sufficient condition on when the chi-squared approximation for the likelihood ratio test fails as the data dimension p increases under $H'_{0,k}$.

Theorem 2.3.2. *Suppose $N \geq p+2$. Under $H'_{0,k}$: $\Sigma = \Lambda_k\Lambda_k^T + \Psi_k$, with given Λ_k and Ψ_k, and $k = k_0$, as $N \to \infty$,*

(i) $\sup_{\alpha\in(0,1)} |\Pr\{T' > \chi^2_{f'}(\alpha)\} - \alpha| \to 0$, *if and only if* $\lim_{n\to\infty} p/N^{1/2} = 0$*;*

(ii) $\sup_{\alpha\in(0,1)} |\Pr\{\rho' \times T' > \chi^2_{f'}(\alpha)\} - \alpha| \to 0$, *if and only if* $\lim_{n\to\infty} p/N^{2/3} = 0$.

Remark II.3. *For the more general testing problem $H_{0,k}$, we need to obtain the maximum likelihood estimators $\hat{\Lambda}_k$ and $\hat{\Psi}_k$, and then conduct the likelihood ratio test with* *chi-squared approximations* (2.11) *or* (2.12). *When the number of latent factors k is fixed compared to N and p, we note that ρ_k/ρ' and f_k/f' asymptotically converge to 1. Furthermore, if $\hat{\Lambda}_k\hat{\Lambda}_k^\top + \hat{\Psi}_k$ approximates the true $\Lambda_k\Lambda_k^\top + \Psi_k$ sufficiently well, we expect that the conclusions in Theorem 2.3.2 would hold for the likelihood ratio test under the null hypothesis $H_{0,k}$ similarly. In particular, when k is fixed as $N \to \infty$, consistent estimation of Λ_k and Ψ_k has been discussed under both fixed p in the classical literature (see, e.g., Anderson, 2003, Theorem 14.3.1) and $p \to \infty$ in recent literature on high-dimensional factor analysis model (see, e.g., Bai and Li, 2012). When k also diverges with N and p, an asymptotic regime that is less investigated in the literature, deriving a similar condition for the chi-squared approximation would require accurate characterizations of the biases of estimating Λ_k and Ψ_k, which, however, would be challenging and need new developments of high-dimensional theory and methodology.*

We next demonstrate the theoretical results by the following numerical study.

Numerical Example 3. *We consider the likelihood ratio test under $H_{0,k}$ with $k = k_0 \in \{1, 3\}$. (I) When $k_0 = 1$, under $H_{0,1}$, we set $\Lambda = \rho \times \mathbf{1}_p$ and $\Psi = (1-\rho^2)\mathrm{I}_p$, with $\rho = 0.3$. (II) When $k_0 = 3$, under $H_{0,3}$, we set $\Psi = (1-\rho^2)\mathrm{I}_p$ and*

$$\Lambda = \begin{bmatrix} \rho \times \mathbf{1}_{p_1} & \mathbf{0}_{p_1} & \mathbf{0}_{p_1} \\ \mathbf{0}_{p_1} & \rho \times \mathbf{1}_{p_1} & \mathbf{0}_{p_1} \\ \mathbf{0}_{p-2p_1} & \mathbf{0}_{p-2p_1} & \rho \times \mathbf{1}_{p-2p_1} \end{bmatrix},$$

where $p_1 = \lfloor p/3 \rfloor$, $\rho = 0.6$, and $\mathbf{1}_{p_1}$ represents a p_1-dimensional vector with all one entries. For both cases, we set $p = \lfloor N^{\varepsilon} \rfloor$, where $N \in \{100, 500, 1000, 2000\}$ and $\varepsilon \in \{8/24, 7/24, ..., 23/24\}$. And we generate each observation $X_i, i = 1, \ldots, N$, from $\mathcal{N}(\mathbf{0}, \Lambda\Lambda^\top + \Psi)$ independently, and conduct the likelihood ratio test with the function factanal() *in* R. *Similarly to Figure II.9, we plot the estimated type I error rates (based on 1000 replications) versus ε for two approximations* (2.11) *and* (2.12), *where the* *results of case (I) are in Figure II.10, and the results of case (II) are in Figure II.11.*

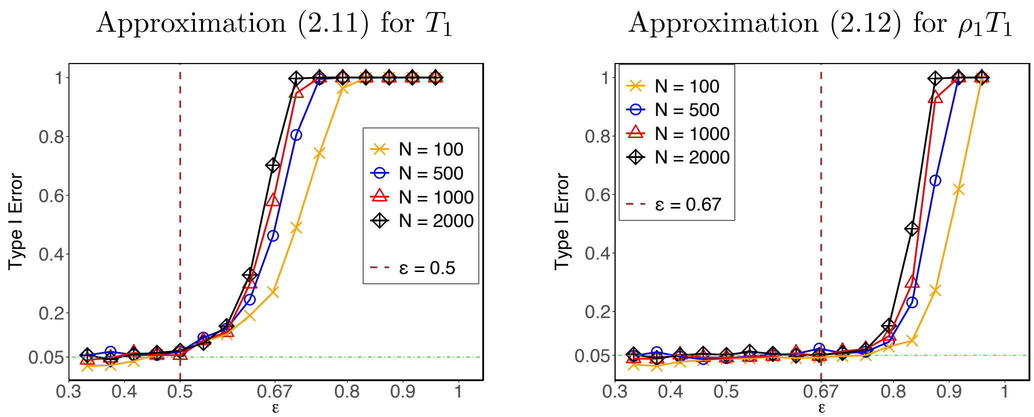


Figure II.10: Empirical type I errors versus ε under Case (I) $k_0 = 1$ in the numerical example 3.

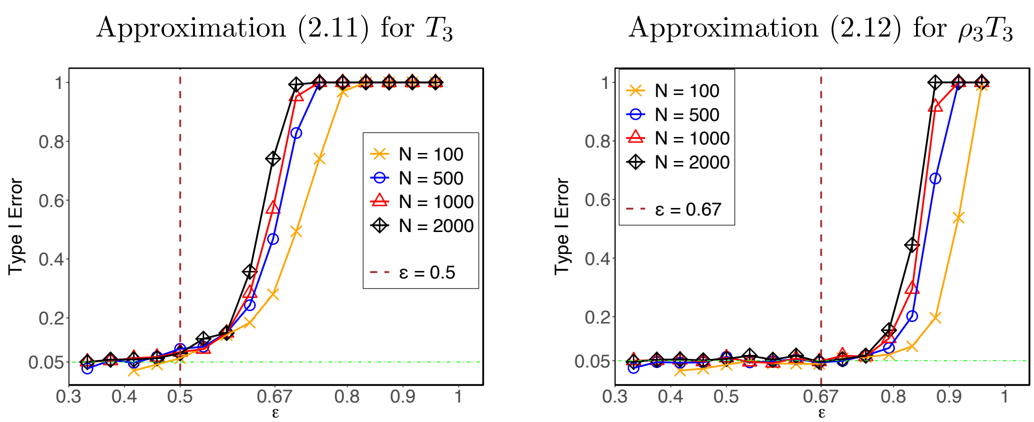


Figure II.11: Empirical type I error versus ε under Case (II) $k_0 = 3$ in the numerical example 3.

Similarly to Numerical Example 2, Numerical Example 3 also demonstrates that the empirical values of ε, where the chi-squared approximations start to fail, are consistent with the corresponding theoretical results. The necessary and sufficient conditions therefore would provide simple quantitative guidelines to check in practice. In addition, it is worth mentioning that the conditions in Theorems 2.3.1 and 2.3.2 also reflect the biases of the chi-squared approximations. For instance, considering the likelihood ratio test for $H_{0,0}$, by the proof of Theorem 2.3.1, when $p/N \to 0$, we obtain that $\mathrm{E}(T_0 - \chi^2_{f_0}) \times \{\mathrm{var}(\chi^2_{f_0})\}^{-1/2}$ is approximately $C_1 p^2/N$, and $\mathrm{E}(\rho_0 \times T_0 - \chi^2_{f_0}) \times \{\mathrm{var}(\chi^2_{f_0})\}^{-1/2}$ is approximately $C_2 p^3/N^2$, where C_1 and C_2 are positive constants. This suggests that the mean of the chi-squared limit will become smaller than the means of T_0 and $\rho_0 T_0$ as p increases, which is consistent with the observed phenomenon in Figure II.8.

2.3.3 Connection with Overestimating the Number of Factors

Figures II.9–II.11 show that the estimated type I error of the likelihood ratio test increases as ϵ increases. This can provide one possible explanation for the well-known finding that the likelihood ratio test tends to overestimate the number of factors in the literature of factor analysis (Hayashi et al., 2007). In particular, let $\hat{k}$ denote the number of factors estimated by the sequential procedure described in Section 2.3.1, and let k_0 denote the true number of factors. Note that in the sequential procedure, rejecting H_{0,k_0} leads to an overestimation of the number of factors, i.e., $\hat{k} > k_0$. Thus, when the type I error of testing H_{0,k_0} inflates as in Figures II.9–II.11, the probability of rejecting H_{0,k_0} would also increase, which consequently suggests an inflation of the probability of overestimating the number of factors, $\hat{k} > k_0$. We next conduct simulation studies to demonstrate the performance of estimating the number of factors using the likelihood ratio test. The numerical results are consistent with the above theoretical analyses and show that the procedure begins to overestimate the number of factors when the type I error begins to inflate.

In particular, consider the simulation setting similar to that in Numerical Example 3, where we take the true number of factors $k_0 \in \{1, 3\}$, sample size $N \in \{500, 1000\}$ and data dimension $p = \lfloor N^\epsilon \rfloor$ for different ϵ values. When conducting the likelihood ratio tests in the sequential procedure, the nominal significance level is set as $\alpha = 0.05$. For each combination of (k_0, N), we use the sequential procedure to estimate the number of factors, denoted as $\hat{k}$. We repeat the procedure 1000 times and estimate the proportions of correct estimation ($\hat{k} = k_0$) and overestimation ($\hat{k} > k_0$), respectively. We present the results for $k_0 = 1, 3$ in Figures II.12 and II.13, respectively, where the results based on the likelihood ratio test without and with the Bartlett correction are given in the left and right columns, respectively.

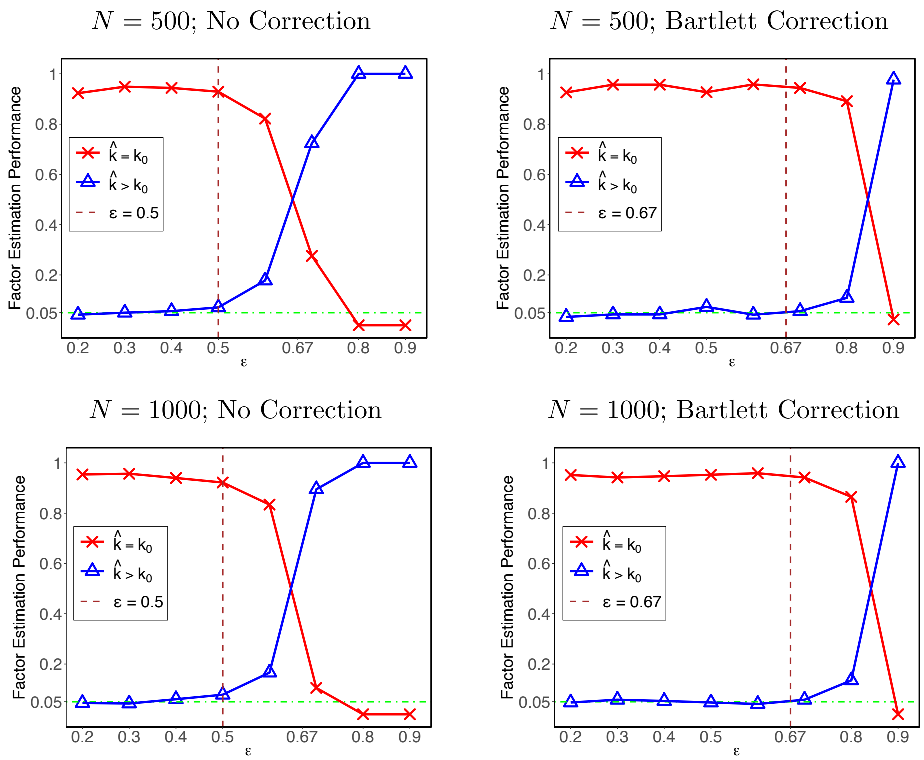


Figure II.12: Performance of the likelihood ratio test for estimating the number of factors when $k_0 = 1$: correct estimation ($\hat{k} = k_0$) and overestimation ($\hat{k} > k_0$).

The numerical results in Figures II.12 and II.13 show that (1) using the likelihood ratio test, the procedure begins to overestimate the number of factors when ϵ approaches 1/2; (2) using the likelihood ratio test with the Bartlett correction, the procedure begins to overestimate the number of factors when ϵ approaches 2/3. These observations, compared with Figures II.9–II.11, suggest that the sequential procedure begins to overestimate the number of factors when the corresponding type I error begins to inflate, which is consistent with our discussions above. Moreover, in Figures II.12 and II.13, when ϵ is small and does not pass the corresponding phase transition boundary, the proportion of overestimation ($\hat{k} > k_0$) is around 0.05. This is because

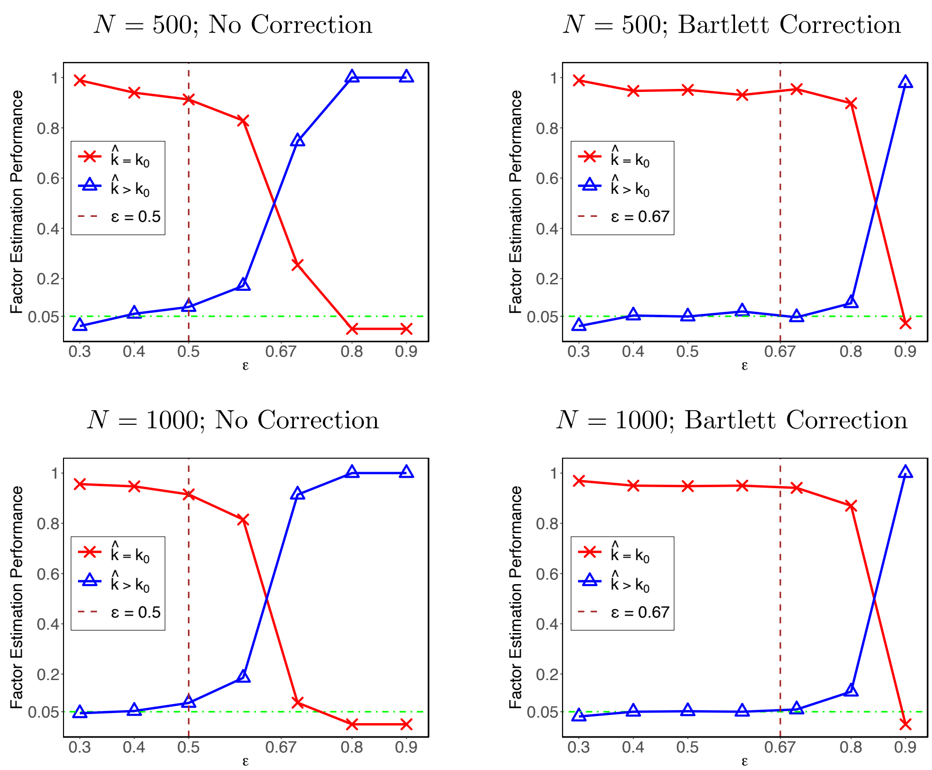


Figure II.13: Performance of the likelihood ratio test for estimating the number of factors when $k_0 = 3$: correct estimation ($\hat{k} = k_0$) and overestimation ($\hat{k} > k_0$).

that rejecting H_{0,k_0} suggests $\hat{k} > k_0$, and the probability of rejecting H_{0,k_0} (type I error of testing H_{0,k_0}) can be asymptotically controlled at the level $\alpha = 0.05$ under the asymptotic regimes derived in Theorems 2.3.1 and 2.3.2.

2.4 Results for Multivariate Linear Regression

2.4.1 Likelihood Ratio Test in Multivariate Linear Regression

Multivariate linear regressions are widely used in econometrics, financial engineering, psychometrics and many other areas of applications to model the relationships between multiple related responses and a set of predictors. Suppose we have n observations of m-dimensional responses $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,m})^\intercal$ and p-dimensional predictors $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^\intercal$, for $i = 1, \ldots, n$. Let $Y = (\mathbf{y}_1, \ldots, \mathbf{y}_n)^\intercal$ be the $n \times m$ response matrix and $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\intercal$ be the $n \times p$ design matrix. The multivariate linear regression model assumes $Y = XB + E$, where B is a $p \times m$ matrix of unknown regression parameters and $E = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n)^\intercal$ is an $n \times m$ matrix of regression errors, with $\boldsymbol{\epsilon}_i$'s independently sampled from an m-dimensional Gaussian distribution $\mathcal{N}(\mathbf{0}, \Sigma)$.

Under the multivariate linear regression model, we are interested in testing the null hypothesis $H_0 : CB = \mathbf{0}_{r\times m}$, where C is an $r \times p$ matrix with rank $r \leq p$ and $\mathbf{0}_{r\times m}$ is an all-zero matrix of size $r \times m$. This is often called general linear hypothesis in multivariate analysis and has been popularly used in multivariate analysis of variance (see, e.g., Muirhead, 2009). Different choices of the testing matrix C are of interest in various applications. For instance, if B is partitioned as $B^\intercal = [B_1^\intercal, B_2^\intercal]$, where B_1 is an $r \times m$ matrix; then the null hypothesis of $B_1 = \mathbf{0}_{r\times m}$ is equivalent to taking $C = [I_r, \mathbf{0}_{r\times(p-r)}]$, which can be used to test the significance of the first r predictors of X. Another example is to test the equivalence of the effects of a set of $r+1$ predictors (such as different levels of some categorical variables), where $C = [I_r, \mathbf{0}_{r\times(p-r-1)}, -\mathbf{1}_r]$ and $\mathbf{1}_r$ represents an all 1 vector of length r.

To test $H_0 : CB = \mathbf{0}_{r\times m}$, a popularly used approach in the literature is the likelihood ratio test (LRT) (Anderson, 2003; Muirhead, 2009). Specifically, when $n > m + p$, Σ is positive definite and X has rank p, the LRT statistic is $L_n = \det(S_E)^{n/2}/\{\det(S_E + S_X)^{n/2}\}$, where $S_E = Y^\intercal[I - X(X^\intercal X)^{-1}X^\intercal]Y$ and $S_X = (C\hat{B})^\intercal[C(X^\intercal X)^{-1}C^\intercal]^{-1}C\hat{B}$ are the residual sum of squares and the regression sum of squares matrices respectively, and $\hat{B} = (X^\intercal X)^{-1}X^\intercal Y$ is the least squares estimator. Assuming m and p are fixed, it is well known that $-2\log L_n$ converges weakly to a χ^2 distribution as $n \to \infty$ under the null hypothesis (Anderson, 2003).

However, in the high-dimensional settings where the dimension parameters (p, m, r) are allowed to increase with n, the LRT suffers from several issues. First, under the null hypothesis, the limiting distribution of $-2\log L_n$ may not be a χ^2 distribution any more. The failure of the χ^2 approximations of LRT distributions under high dimensions has been studied by researchers under various model settings. For instance, Bai et al. (2009) examined two LRTs on testing covariance matrices, showed that their χ^2 approximations perform poorly, and proposed the corrected normal limiting distributions. Jiang and Yang (2013) and Jiang and Qi (2015) studied classical LRTs on testing sample means and covariance matrices, and showed that the χ^2 approximations also fail as the dimensions increase. Moreover, Bai et al. (2013) considered the LRT on testing linear hypotheses in high-dimensional multivariate linear regressions, demonstrated the failure of χ^2 approximation and derived the corrected LRT. Note that Bai et al. (2013) only considered the high-dimensional settings where m, r and $n - p$ are proportional to each other with $m \le r$. Despite these existing works, it is still unclear under which asymptotic regimes the χ^2 approximation of LRT starts to fail. An answer to this question would provide insights for practitioners especially when analyzing data with m/n and p/n small but not negligible.

To address this issue, we derive the asymptotic boundary when the χ^2 approximation fails as the dimension parameters (p, m, r) increase with the sample size

n. Moreover, we develop the corrected limiting distribution of $\log L_n$ for a general asymptotic regime of (p, m, r, n).

2.4.2 Phase Transition Boundary

In traditional multivariate regression analysis where the dimension parameters (p, m, r) are considered as fixed numbers, the χ^2 approximation of the LRT,

$$-2 \log L_n \xrightarrow{D} \chi^2_{mr}, \qquad \text{as } n \to \infty, \tag{2.13}$$

is used for $H_0 : CB = \mathbf{0}_{r\times m}$ (Muirhead, 2009; Anderson, 2003), where $\xrightarrow{D}$ denotes the convergence in distribution. However, it has been noted that the χ^2 approximation of the distribution of the LRT often performs poorly in high-dimensional applications (see, e.g., Bai and Saranadasa, 1996; Jiang et al., 2012; Bai et al., 2009, 2013; Jiang and Yang, 2013).

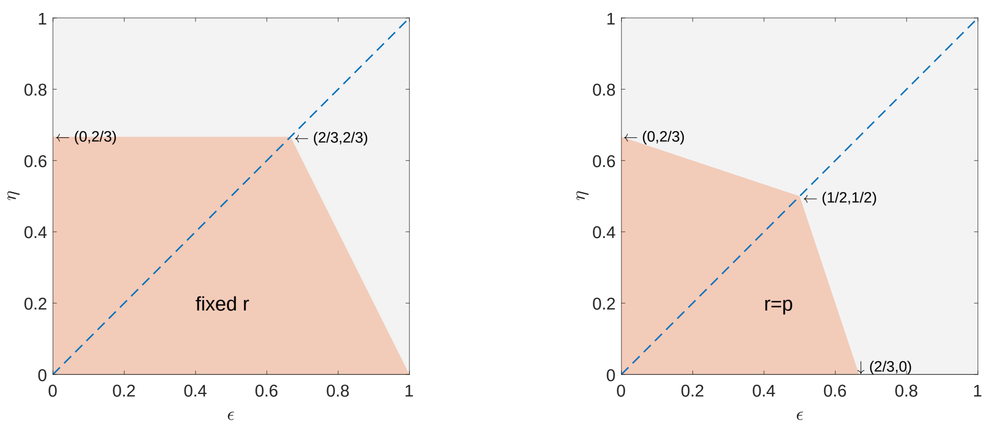
As the three dimension parameters (m, p, r) are allowed to grow with n, it is of interest to examine the phase transition boundary where the χ^2 approximation fails. This is described in the following theorem.

Theorem 2.4.1 (Without the Bartlett Correction)**.** *Consider $n > p + m$ and $p \geq r$. Let $\chi^2_{mr}(\alpha)$ denote the upper α-quantile of χ^2_{mr} distribution.*
(i) When $mr \to \infty$ and $\max\{p, m, r\}/n \to 0$ as $n \to \infty$, $P\{-2\log L_n > \chi^2_{mr}(\alpha)\} \to \alpha$, for any significance level α, if and only if

$$\lim_{n\to\infty} \sqrt{mr}(p + m/2 - r/2)n^{-1} = 0. \tag{2.14}$$

(ii) When mr is finite, $P\{-2\log L_n > \chi^2_{mr}(\alpha)\} \to \alpha$, if and only if $\lim_{n\to\infty} p/n = 0$.

Theorem 2.4.1 gives the sufficient and necessary condition on (m, p, r, n) such that the χ^2 approximation (2.13) fails. We note that although (2.14) is obtained when $mr \to \infty$, (2.14) becomes $\lim_{n\to\infty} p/n = 0$ with finite m and r, in agreement with the conclusion when mr is finite. To further examine the implications of (2.14), we consider two special cases. Specifically, let $m = \lfloor n^\eta \rfloor$ and $p = \lfloor n^\epsilon \rfloor$ with η and $\epsilon \in (0,1)$, where $\lfloor \cdot \rfloor$ denotes the floor of a number. When r is fixed, (2.14) implies $\sqrt{m}(p + m/2) = o(n)$, that is, $\max\{\epsilon, \eta\} + \eta/2 < 1$. When $r = p = \lfloor n^\epsilon \rfloor$, (2.14) implies $\sqrt{mp}(p+m) = o(n)$, that is, $\max\{\epsilon, \eta\} + (\eta + \epsilon)/2 < 1$. For these two cases, we correspondingly give two (η, ϵ)-regions in Figure II.14 satisfying the constraint (2.14). In these two regions, when ϵ becomes close to 0, the largest η approaches 2/3. This implies that when p is small, the largest m such that (2.14) holds is of order $n^{2/3}$. It is the same for both fixed r and $r = p$ cases as p is small and $r \le p$. In addition, when η goes to 0, the largest ϵ values under fixed r and $r = p$ cases converge to 1 and 2/3 respectively. This indicates that when m is small, the largest p values satisfying (2.14) are of order n and $n^{2/3}$ for the two cases respectively. Moreover, when $m = p$, the largest orders of m and p for the two cases are $n^{2/3}$ and $n^{1/2}$ respectively.



(a) r is fixed (b) $r = p$

Figure II.14: Illustration of phase transition boundaries in (2.14): η versus ϵ when $m = \lfloor n^\eta \rfloor$ and $p = \lfloor n^\epsilon \rfloor$. Scenario (a): r is fixed; Scenario (b): $r = p$.

To illustrate this phase transition phenomenon, we present a simple simulation experiment. We set $\Sigma = I_m$, and estimate the type I errors of the χ^2 approximation (2.13) with 10^4 repetitions under the following four cases: (a) fixed $m = r =$

2 and $p = \lfloor n^\eta \rfloor$; (b) fixed $p = r = 2$ and $m = \lfloor n^\eta \rfloor$; (c) fixed $m = 2$ and $p = r = \lfloor n^\eta \rfloor$; (d) $p = m = r = \lfloor n^\eta \rfloor$, where $\eta \in \{1/24, \dots, 23/24\}$. In Figure II.15, we plot the estimated type I errors against η values for $n = 100$ and 300 respectively. The plots show consistent patterns with the theoretical results. In particular, when $p = m = r = \lfloor n^\eta \rfloor$, the χ^2 approximation begins to fail for η around 1/2. When p and r are fixed and $m = \lfloor n^\eta \rfloor$, or when m is fixed and $p = r = \lfloor n^\eta \rfloor$, the χ^2 approximation begins to fail for η around 2/3. When m and r are fixed and $p = \lfloor n^\eta \rfloor$, the χ^2 approximation begins to fail for η larger than the other three cases, which is consistent with the theoretical results.

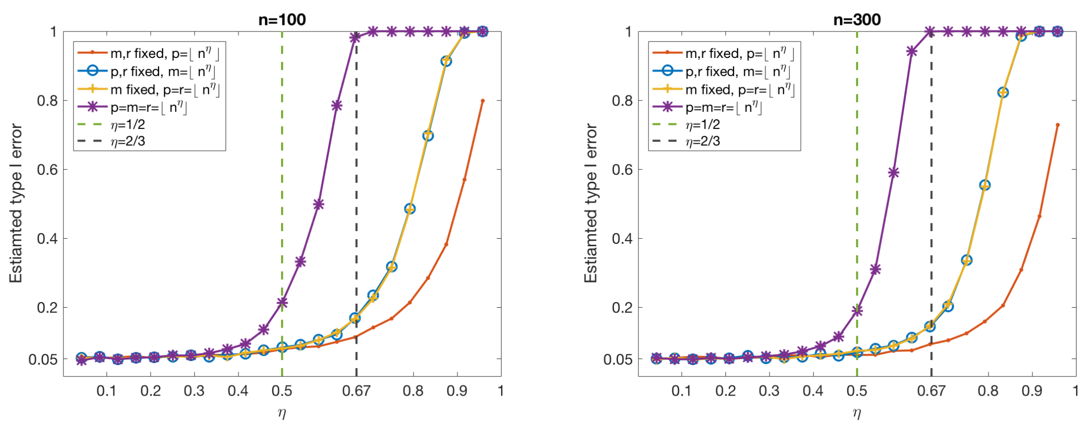


Figure II.15: Empirical type I errors versus η using χ^2 approximation (2.13) when $n = 100$ and $n = 300$.

It is worth mentioning that the sufficient and necessary constraint (2.14) also characterizes the bias of the χ^2 approximation. Specifically, under the conditions of Theorem 2.4.1, $\mathrm{E}(-2\log L_n - \chi^2_{mr})/\sqrt{\mathrm{var}(\chi^2_{mr})} = \sqrt{mr}(p + m/2 - r/2 + 1/2)n^{-1}\{1 + o(1)\}$. Thus when (p, m, r) are large such that (2.14) is violated and the χ^2 approximation fails, the bias of the χ^2 approximation increases as $\sqrt{mr}(p + m/2 - r/2 + 1/2)n^{-1}$ increases.

In the classic regime with fixed m and p, researchers have also proposed the Bartlett correction of the LRT that $-2\rho \log L_n \xrightarrow{D} \chi^2_{mr}$, where $\rho = 1 - (p - r/2 + m/2 + 1/2)/n$. In particular, for any $z \in \mathbb{R}$, this corrected approximation gets rid of the first order approximation error $O(n^{-1})$; that is, for any z, $P(-2\rho \log L_n < z) - P(\chi^2_{mr} < z) = O(n^{-2})$ when m and p are fixed. Similarly to Theorem 1, the χ^2 approximation with Bartlett correction also fails as m and p increase with n. The phase transition boundary is characterized in the following result.

Theorem 2.4.2 (With the Bartlett Correction)**.** *Consider $n > p + m$ and $p \geq r$.*

(i) When $mr \to \infty$ and $\max\{p, m, r\}/n \to 0$ as $n \to \infty$, for any significance level α, $P\{-2\rho \log L_n > \chi^2_{mr}(\alpha)\} \to \alpha$, if and only if $\lim_{n\to\infty} \sqrt{mr}(r^2 + m^2)n^{-2} = 0$.

(ii) When mr is finite, $P\{-2\rho \log L_n > \chi^2_{mr}(\alpha)\} \to \alpha$, if and only if $n - p \to \infty$.

Theorem 2.4.2 suggests that when m and r are fixed, the corrected LRT approximation holds when $n - p \to \infty$. When $mr \to \infty$, the phase transition threshold in Theorem 2.4.2 only involves m and r. In particular, when r is fixed and $m = \lfloor n^\eta \rfloor$, or when m is fixed and $r = \lfloor n^\eta \rfloor$, the χ^2 approximation with Bartlett correction fails when $\eta \geq 4/5$; when $m = r = \lfloor n^\eta \rfloor$, the corrected approximation fails when $\eta \geq 2/3$.

To illustrate the phenomenon, we also present a numerical experiment on the χ^2 approximation with Bartlett correction in Figure II.16 under the same set-up as in Figure II.15. It shows that when m and r are fixed and $p = \lfloor n^\eta \rfloor$, the type I errors are well controlled for large η approaching 1. Moreover, when p and r are fixed and $m = \lfloor n^\eta \rfloor$, or when m is fixed and $p = r = \lfloor n^\eta \rfloor$, the corrected χ^2 approximation begins to fail around $\eta = 4/5$. When $p = m = r = \lfloor n^\eta \rfloor$, the corrected χ^2 approximation begins to fail around $\eta = 2/3$. These numerical results are also consistent with the theory.

2.4.3 Alternative High-Dimensional Limit

More generally, to have a unified limiting distribution for analyzing high-dimensional data under a general asymptotic region of (m, p, r, n), we derive a corrected normal limiting distribution for the LRT statistic.

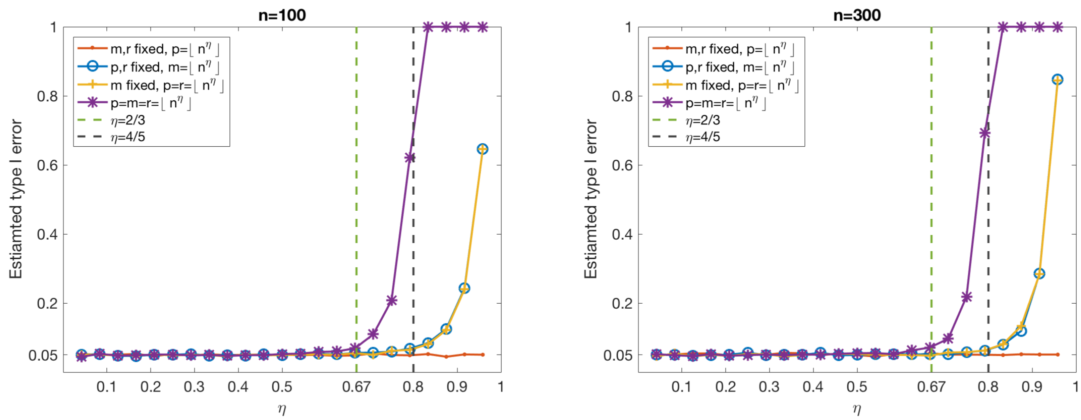


Figure II.16: Empirical type I errors versus η using the χ^2 approximation with the Bartlett correction when $n = 100$ and $n = 300$.

Theorem 2.4.3. *When $n > p+m$, $p \geq r$, $mr \to \infty$, and $n-p-\max\{m-r, 0\} \to \infty$ as $n \to \infty$, the LRT statistic L_n has corrected form T_1 satisfying*

$$T_1 := \frac{-2\log L_n + \mu_n}{n\sigma_n} \xrightarrow{D} \mathcal{N}(0,1), \tag{2.15}$$

where $\sigma_n^2 = 2\log(n+r-p-m)(n-p) - 2\log(n-p-m)(n+r-p)$, and

$$\mu_n = n(n-m-p-1/2)\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)} + nr\log\frac{(n+r-p-m)}{(n+r-p)} + nm\log\frac{(n-p)}{(n+r-p)}.$$

The theorem above covers the asymptotic regime where $mr \to \infty$, $\max\{p, m, r\}/n \to 0$ and the constraint (2.14) holds; under this region, we can show that $\mu_n \to -mr$ and $(n\sigma_n)^2 \to 2mr$, which are consistent with the mean and variance of χ^2_{mr} approximation. In addition, although Theorem 2.4.3 requires $mr \to \infty$, the normal approximation (2.15) could still perform well when m or r is small, as long as mr is large enough.

Alternatively, under some high dimensional settings, we can check that no χ^2 or even noncentral χ^2 distribution could match the asymptotic mean and variance of

$-2\log L_n$ in Theorem 2.4.3. Specifically, if the distribution of $-2\log L_n$ could be approximated by some χ^2 distribution, then we should have $-(n\sigma_n)^2/\mu_n \to 2$, which is, however, not satisfied as $p/n, m/n$ and r/n increase; or if the distribution of $-2\log L_n$ could be approximated by some noncentral χ^2 distribution with degrees of freedom k_n, then we should have $k_n = -2\mu_n - n^2\sigma_n^2/2$, which, however, can become negative as $p/n, m/n$ and r/n increase. Thus it implies that the χ^2-type approximation for $-2\log L_n$ could fail fundamentally under high dimensions.

Remark II.4. *A similar result on the asymptotic normality of* $\log L_n$ *in Theorem 2.4.3 was proved in Zheng (2012) and Bai et al. (2013). However, there are several differences between our result and theirs. First, our asymptotic regime is more general. Specifically, Zheng (2012) and Bai et al. (2013) requires that* $m < r$, $\min\{m, r\} \to \infty$, *and* $m/(n-p)$ *converges to a constant in* $(0, 1)$, *while we only need* $mr \to \infty$ *and* $n - p - \max\{m - r, 0\} \to \infty$. *Our analysis covers the case when* $m/(n-p) \to 0$ *and even when the limit does not exists. Second, the proofs of Zheng (2012) and Bai et al. (2013) are based on the random matrix theory, while we prove Theorem 2.4.3 by a moment generating function technique motivated by Jiang and Yang (2013).*

CHAPTER III

Adaptively Powerful High-Dimensional Tests

3.1 Introduction

Global Hypothesis Testing and Existing Methods In many large-scale inference problems, one is often interested in globally testing some overall patterns of low-dimensional features of the high-dimensional random observations. One example is genome-wide association studies (GWAS), whose primary goal is to identify single nucleotide polymorphisms (SNPs) associated with certain complex diseases of interest. A popular approach in GWAS is to perform univariate tests which examine each SNP one by one. This however may lead to low statistical power due to the weak effect size of each SNP (Manolio et al., 2009) and the small statistical significance threshold ($\sim 10^{-8}$) chosen to control the multiple-comparison type I error (Kim et al., 2016). Researchers therefore have proposed to globally test a genetic marker set with many SNPs (Wang et al., 2011; Kim et al., 2016) in order to achieve higher statistical power and to better understand the underlying genetic mechanisms.

In this section, we focus on a family of global testing problems in the high-dimensional setting, including testing of mean vectors, covariance matrices and regression coefficients in generalized linear models. These problems can be formulated as testing $H_0 : \mathcal{E} = \mathbf{0}$, where $\mathbf{0}$ is an all zero vector, $\mathcal{E} = \{e_l : l \in \mathcal{L}\}$ is a parameter vector with $\mathcal{L}$ being the index set, and e_l's being the corresponding parameters of interest, e.g., elements in mean vectors, covariance matrices or coefficients in generalized linear models. For the global testing problem $H_0 : \mathcal{E} = \mathbf{0}$ versus $H_A : \mathcal{E} \neq \mathbf{0}$, two different types of methods are often used in the literature. One is sum-of-squares-type statistics. They are usually powerful against "dense" alternatives, where $\mathcal{E}$ has a high proportion of nonzero elements with a large $\|\mathcal{E}\|_2 = \sum_{l\in\mathcal{L}} e_l^2$ or its weighted variants. See examples in mean testing (e.g., Bai and Saranadasa, 1996; Goeman et al., 2006; Srivastava and Du, 2008; Chen and Qin, 2010; Chen et al., 2019a; Gregory et al., 2015; Srivastava et al., 2016) and covariance testing (e.g., Bai et al., 2009; Ledoit and Wolf, 2002; Chen et al., 2010; Li and Chen, 2012). The other is maximum-type statistics. They are usually powerful against "sparse" alternatives, where $\mathcal{E}$ has few nonzero elements with a large $\|\mathcal{E}\|_\infty$ (e.g., Jiang, 2004; Liu et al., 2008; Hall and Jin, 2010; Cai and Jiang, 2011; Cai et al., 2013, 2014; Shao and Zhou, 2014). More recently, Fan et al. (2015); Yang and Pan (2017) also proposed to combine these two kinds of test statistics. However, for denser or only moderately dense alternatives, neither of these two types of statistics may be powerful, as will be further illustrated in this chapter both theoretically and numerically. Importantly, in real applications, the underlying truth is usually unknown, which could be either sparse, dense, or in-between. As the global testing could be highly underpowered if an unsuitable testing method is used (e.g., Colantuoni et al., 2011), it is desired in practice to have a testing procedure with high statistical power against a variety of alternatives.

A Family of Asymptotically Independent U-Statistics To address these issues, we propose a U-statistics framework and introduce its applications to adaptive high-dimensional testing. The U-statistics framework constructs *unbiased* and *asymptotically independent* estimators of $\|\mathcal{E}\|_a^a := \sum_{l\in\mathcal{L}} e_l^a$ for different (positive) integers a, where $a = 2$ corresponds to a sum-of-squares-type statistic, and an even integer $a \to \infty$ yields a maximum-type statistic. The adaptive testing then combines the information from different $\|\mathcal{E}\|_a^a$'s, and our power analysis shows that it is powerful against a wide range of alternatives, from highly sparse, moderately sparse to dense, to highly dense.

To illustrate our idea, suppose $\mathbf{z}_1, \ldots, \mathbf{z}_n$ are n independent and identically distributed (i.i.d.) copies of a random vector $\mathbf{z}$. We consider the setting where each parameter e_l has an unbiased kernel function estimator $K_l(\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{\gamma_l}})$, and γ_l is the smallest integer such that for any $1 \le i_1 \ne \ldots \ne i_{\gamma_l} \le n$, $\mathrm{E}[K_l(\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{\gamma_l}})] = e_l$. This includes many testing problems on moments of low orders, such as entries in mean vectors, covariance matrices and score vectors of generalized linear models, which shall be discussed in details. The family of U-statistics can be constructed generally as follows. For integers $a \ge 1$, and $1 \le i_1 \ne \ldots \ne i_{\gamma_l} \ne \ldots \ne i_{(a-1)\times\gamma_l+1} \ldots \ne i_{a\times\gamma_l} \le n$, since the $\mathbf{z}$'s are i.i.d., we have $\mathrm{E}[K_l(\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_{\gamma_l}}) \cdots K_l(\mathbf{z}_{i_{(a-1)\times\gamma_l+1}}, \ldots, \mathbf{z}_{i_{a\times\gamma_l}})] = e_l^a$. Therefore, we can construct an unbiased estimator of the parameters of augmented powers e_l^a with different a. Then $\|\mathcal{E}\|_a^a$ has an unbiased estimator

$$\mathcal{U}(a) = \sum_{l\in\mathcal{L}} (P^n_{a\times\gamma_l})^{-1} \sum_{1\le i_1\ne\ldots\ne i_{a\times\gamma_l}\le n} \prod_{k=1}^{a} K_l(\mathbf{z}_{i_{(k-1)\times\gamma_l+1}}, \ldots, \mathbf{z}_{i_{k\times\gamma_l}}), \tag{3.1}$$

where $P^n_k = n!/(n-k)!$ denotes the number of k-permutations of n. We call a the order of the U-statistic $\mathcal{U}(a)$. If $a > b$, we say $\mathcal{U}(a)$ is of higher order than $\mathcal{U}(b)$ and vice versa.

This construction procedure can be applied to many testing problems. We give three common examples below for illustration and more detailed case-studies will be discussed in the following sections.

Example III.1. *Consider one-sample mean testing of* $H_0 : \boldsymbol{\mu} = \mathbf{0}$*, where* $\mathcal{E} = \boldsymbol{\mu}$ *is the mean vector of a* p*-dimensional random vector* $\mathbf{x}$*. Suppose* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *are* n *i.i.d. copies of* $\mathbf{x}$*. For each* $i = 1, \ldots, n$*,* $j = 1, \ldots, p$*,* $x_{i,j}$ *is a simple unbiased estimator of* μ_j*, then we can take the kernel function* $K_j(\mathbf{x}_i) = x_{i,j}$*. Following* (3.1)*, we know* *the U-statistic* $\mathcal{U}(a) = (P^n_a)^{-1} \sum_{j=1}^p \sum_{1\leq i_1\neq\ldots\neq i_a\leq n} \prod_{k=1}^a x_{i_k,j}$ *is an unbiased estimator of* $\|\mathcal{E}\|_a^a = \|\boldsymbol{\mu}\|_a^a = \sum_{j=1}^p \mu_j^a$. *Please see Section 3.3 for the two-sample mean testing example and related theoretical properties.*

Example III.2. *Suppose* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *are* n *i.i.d. copies of a random vector* $\mathbf{x}$ *with mean vector* $\boldsymbol{\mu} = \mathbf{0}$ *and covariance matrix* $\boldsymbol{\Sigma} = \{\sigma_{j_1,j_2}\}_{p\times p}$. *For covariance testing* $H_0 : \sigma_{j_1,j_2} = 0$ *for any* $1 \leq j_1 \neq j_2 \leq p$, *we have* $\mathcal{E} = \{\sigma_l : l \in \mathcal{L}\}$ *with* $\mathcal{L} = \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p\}$. *Since* $x_{i,j_1}x_{i,j_2}$ *is a simple unbiased estimator of* σ_{j_1,j_2}, *then for each pair* $l = (j_1, j_2) \in \mathcal{L}$, *we can take the kernel function* $K_l(\mathbf{x}_i) = x_{i,j_1}x_{i,j_2}$. *Following* (3.1), *the U-statistic* $\mathcal{U}(a) = (P^n_a)^{-1} \sum_{1\leq j_1\neq j_2\leq p} \sum_{1\leq i_1\neq\ldots\neq i_a\leq n} \prod_{k=1}^a (x_{i_k,j_1}x_{i_k,j_2})$ *is an unbiased estimator of* $\|\mathcal{E}\|_a^a = \sum_{1\leq j_1\neq j_2\leq p} \sigma^a_{j_1,j_2}$. *Please see Section 3.2 for the general case with unknown* $\boldsymbol{\mu}$.

Example III.3. *Consider a response variable* y *and its covariates* $\mathbf{x} \in \mathbb{R}^p$ *following a generalized linear model:* $\mathrm{E}(y|\mathbf{x}) = g^{-1}(\mathbf{x}^\intercal\boldsymbol{\beta})$, *where* g *is the canonical link function and* $\boldsymbol{\beta} \in \mathbb{R}^p$ *are the regression coefficients. Suppose that* $(\mathbf{x}_i, y_i)$, $i = 1, \ldots, n$, *are i.i.d. copies of* $(\mathbf{x}, y)$. *To test* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ *versus* $H_A : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, *the score vectors* $(S_{i,j} = (y_i - \mu_{0,i})x_{i,j} : j = 1, \ldots, p)^\intercal$ *are often used in the literature, where* $\mu_{0,i} = g^{-1}(\mathbf{x}_i^\intercal\boldsymbol{\beta}_0)$. *Note that* $\mathrm{E}(S_{i,j}) = 0$ *under* H_0. *Thus to test* H_0, *we can take* $\mathcal{E} = \{\mathrm{E}(S_{i,j}) : j = 1, \ldots, p\}$ *and use the U-statistic* $\mathcal{U}(a) = (P^n_a)^{-1} \sum_{j=1}^p \sum_{1\leq i_1\neq\ldots\neq i_a\leq n} \prod_{k=1}^a S_{i_k,j}$, *which is an unbiased estimator of* $\|\mathcal{E}\|_a^a = \sum_{j=1}^p \{\mathrm{E}(S_{i,j})\}^a$. *Please see Section 3.5.*

Related Literature For high-dimensional testing, some other adaptive testing procedures have recently been proposed in the literature (Pan et al., 2014; Xu et al., 2016; Wu et al., 2019). These works combine the p-values of a family of sum-of-powered statistics that are powerful against different $\|\mathcal{E}\|_a^a$'s. However in these existing works, to evaluate the p-value of the adaptive test statistic, the joint asymptotic distribution of the statistics is difficult to obtain or calculate. Accordingly computationally expensive resampling methods are often used in practice (Pan et al., 2014; Kim et al., 2016; Xu et al., 2017). For some special cases such as testing means and the coefficients of generalized linear models, Xu et al. (2016) and Wu et al. (2019) derived the limiting distributions of the test statistics under the framework of a family of von Mises V-statistics. However, the constructed V-statistics are usually *correlated* and *biased* estimators of the target $\|\mathcal{E}\|_a^a$. It follows that in Xu et al. (2016) and Wu et al. (2019), numerical approximations are still needed to calculate the tail probabilities of the adaptive test statistics. In addition, these existing adaptive testing works mainly focus on the first-order moments, and their results do not directly apply to testing second-order moments, such as covariance matrices.

To overcome these issues, this chapter studies the proposed family of unbiased U-statistics. There are some other recent works providing important results on high-dimensional U-statistics (e.g., Chen, 2018; Leung and Drton, 2018; Zhong and Chen, 2011). For instance, Zhong and Chen (2011) considered testing the regression coefficients in linear models using the fourth-order U-statistic; Leung and Drton (2018) studied the limiting distributions of rank-based U-statistics; and Chen (2018) studied bootstrap approximation of the second-order U-statistics. However, these results do not directly apply to the high-order U-statistics considered in this chapter.

Our Contributions We establish the theoretical properties of the U-statistics in various high dimensional testing problems, including testing mean vectors, regression coefficients of generalized linear models, and covariance matrices. Our contributions are summarized as follows.

Under the null hypothesis, we show that the normalized U-statistics of different finite orders are jointly normally distributed. The result applies generally for any asymptotic regime with $n \to \infty$ and $p \to \infty$. In addition, we prove that all the finite-order U-statistics are asymptotically independent with each other under the null hypothesis. Moreover, we prove that U-statistics of finite orders are also asymptotically independent of the maximum-type test statistic with a limiting extreme value distribution.

Under the alternative hypothesis, we further analyze the asymptotic power for U-statistics of different orders. We show that when $\mathcal{E}$ has denser nonzero entries, $\mathcal{U}(a)$'s of lower orders tend to be more powerful; and when $\mathcal{E}$ has sparser nonzero entries, $\mathcal{U}(a)$'s of higher orders tend to be more powerful. More interestingly, we show that in the boundary case of "moderate" sparsity levels, $\mathcal{U}(a)$ with a finite $a > 2$ gives the highest power among the family of U-statistics, clearly indicating the inadequacy of both the sum-of-squares- and the maximum-type statistics.

An important application of the independence property among $\mathcal{U}(a)$'s is to construct adaptive testing procedures by combining the information of different $\mathcal{U}(a)$'s, whose univariate distributions or p-values can be easily combined to form a joint distribution to calculate the p-value of an adaptive test statistic. Compared with other existing works (e.g., Xu et al., 2016; Wu et al., 2019), numerical approximations of tail probabilities are no longer needed. As shown in the power analysis, an adaptive integration of information across different tests leads to a powerful testing procedure.

In the remaining of this chapter, Section 3.2 illustrates the framework by a covariance testing problem, and Sections 3.5–3.5 establish similar results in other high-dimensional testing problems, including testing means, regression coefficients and two-sample covariances.

3.2 Motivating Example: One-Sample Covariance Test

This section illustrates the framework with a motivating example of a one-sample covariance testing problem. We showcase the study of the one-sample covariance testing problem since this is more challenging than mean testing due to the two-way dependency structure, and the one-sample problem can be used as the building block for more general cases. Specifically, we focus on testing

$$H_0 : \sigma_{j_1,j_2} = 0 \quad \forall\ 1 \leq j_1 \neq j_2 \leq p, \tag{3.2}$$

where $\mathbf{\Sigma} = \{\sigma_{j_1,j_2} : 1 \leq j_1, j_2 \leq p\}$ is the covariance matrix of a p-dimensional real-valued random vector $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ with $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$. The observed data include n i.i.d. copies of $\mathbf{x}$, denoted by $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^\intercal$. In factor analysis, testing H_0 in (3.2) can be used to examine whether $\mathbf{\Sigma}$ has any significant factor or not (Anderson, 2003).

Global testing of covariance structure plays an important role in many statistical analysis and applications; see a review in Cai (2017). Conventional tests include the likelihood ratio test, John's test, and Nagao's test, etc. (Anderson, 2003; Muirhead, 2009). These methods, however, often fail in the high-dimensional setting when both $n, p \to \infty$. To address this issue, new procedures have been recently proposed (e.g., Bai et al., 2009; Jiang and Yang, 2013; Johnstone, 2001; Soshnikov, 2002; Schott, 2007; Péché, 2009; Ledoit and Wolf, 2002; Chen et al., 2010; Jiang, 2004; Liu et al., 2008; Cai and Jiang, 2011; Li and Chen, 2012; Shao and Zhou, 2014; Lan et al., 2015). However, these methods might suffer from loss of power when the sparsity level of the alternative covariance matrix varies. In the following subsections, we introduce the U-statistics framework, study their asymptotic properties, and develop a powerful adaptive testing procedure.

3.2.1 Asymptotically Independent U-Statistics

For testing (3.2), the set of parameters that we are interested in is $\mathcal{E} = \{\sigma_{j_1,j_2} : 1 \leq j_1 \neq j_2 \leq p\}$. Following the previous analysis of (3.1), since σ_{j_1,j_2} has a simple unbiased estimator $x_{i_1,j_1}x_{i_1,j_2} - x_{i_1,j_1}x_{i_2,j_2}$ with $1 \leq i_1 \neq i_2 \leq n$, then for integers

$a \geq 1$, an unbiased U-statistic of $\|\mathcal{E}\|_a^a = \sum_{1\leq j_1\neq j_2\leq p} \sigma_{j_1,j_2}^a$ is

$$\mathcal{U}(a) = (P_{2a}^n)^{-1} \sum_{1\leq j_1\neq j_2\leq p} \sum_{1\leq i_1\neq \ldots \neq i_{2a}\leq n} \prod_{k=1}^{a} (x_{i_{2k-1},j_1} x_{i_{2k-1},j_2} - x_{i_{2k-1},j_1} x_{i_{2k},j_2}).$$

This is equivalent to

$$\begin{aligned} \mathcal{U}(a) \quad = \quad & \sum_{1\leq j_1\neq j_2\leq p} \sum_{c=0}^{a} (-1)^c \binom{a}{c} \frac{1}{P_{a+c}^n} \sum_{1\leq i_1\neq \ldots \neq i_{a+c}\leq n} \\ & \prod_{k=1}^{a-c} (x_{i_k,j_1} x_{i_k,j_2}) \prod_{s=a-c+1}^{a} x_{i_s,j_1} \prod_{t=a+1}^{a+c} x_{i_t,j_2}. \end{aligned} \tag{3.3}$$

Remark III.1. *The U-statistics can be constructed by another method equivalently. Given $1 \leq j_1 \neq j_2 \leq p$, define $\varphi_{j_1,j_2} = \sigma_{j_1,j_2} + \mu_{j_1}\mu_{j_2}$. Then*

$$\sum_{1\leq j_1\neq j_2\leq p} \sigma_{j_1,j_2}^a = \sum_{1\leq j_1\neq j_2\leq p} \sum_{c=0}^{a} \binom{a}{c} \varphi_{j_1,j_2}^{a-c} \times (-\mu_{j_1}\mu_{j_2})^c, \tag{3.4}$$

which is a polynomial function of the moments μ_j and φ_{j_1,j_2}. Since μ_j and φ_{j_1,j_2} have unbiased estimators $x_{i,j}$ and $x_{i,j_1}x_{i,j_2}$ respectively, then for $1 \leq i_1 \neq \ldots \neq i_{a+c} \leq n$, $\mathrm{E}(\prod_{k=1}^{a-c} x_{i_k,j_1} x_{i_k,j_2} \prod_{s=a-c+1}^{a} x_{i_s,j_1} \prod_{t=a+1}^{a+c} x_{i_t,j_2}) = \varphi_{j_1,j_2}^{a-c} \mu_{j_1}^c \mu_{j_2}^c$. Given this and (3.4)*, the U-statistics* (3.3) *can be obtained.*

Remark III.2. *The summed term with $c = 0$ in* (3.3) *is*

$$\tilde{\mathcal{U}}(a) := (P_a^n)^{-1} \sum_{1\leq i_1\neq \cdots \neq i_a\leq n} \sum_{1\leq j_1\neq j_2\leq p} \prod_{k=1}^{a} (x_{i_k,j_1} x_{i_k,j_2}), \tag{3.5}$$

which has the same form as the simplified U-statistic for mean zero observations in Example III.2, and is shown to be the leading term of (3.3) *in proof.*

It follows that the constructed U-statistics (3.3) enjoy two nice properties on location invariance and unbiasedness.

Property 3.2.1 (Location Invariance). *$\mathcal{U}(a)$ constructed as in* (3.3) *is location invariant; that is, for any vector $\boldsymbol{\Delta} \in \mathbb{R}^p$, the U-statistic constructed based on the transformed data $\{\mathbf{x}_i + \boldsymbol{\Delta} : i = 1, \dots, n\}$ is still $\mathcal{U}(a)$.*

Property 3.2.2 (Unbiasedness). *For any integer a, $\mathrm{E}[\mathcal{U}(a)] = \sum_{1 \le j_1 \ne j_2 \le p} \sigma^a_{j_1,j_2}$. Under H_0 in* (3.2), $\mathrm{E}[\mathcal{U}(a)] = 0$.

We next study the limiting properties of the constructed U-statistics under H_0 given the following assumptions on the random vector $\mathbf{x} = (x_1, \dots, x_p)^\intercal$.

Condition 3.2.1 (Moment assumption). $\lim_{p\to\infty} \max_{1\le j\le p} \mathrm{E}(x_j - \mu_j)^8 < \infty$ *and* $\lim_{p\to\infty} \min_{1\le j\le p} \mathrm{E}(x_j - \mu_j)^2 > 0$.

Condition 3.2.2 (Dependence assumption). *For a sequence of random variables $\mathbf{z} = \{z_j : j \ge 1\}$ and integers $a < b$, let $\mathcal{Z}_a^b$ be the σ-algebra generated by $\{z_j : j \in \{a, \dots, b\}\}$. For each $s \ge 1$, define the α-mixing coefficient $\alpha_{\mathbf{z}}(s) = \sup_{t\ge1}\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{Z}_1^t, B \in \mathcal{Z}_{t+s}^\infty\}$. We assume that under H_0, $\mathbf{x}$ is α-mixing with $\alpha_{\mathbf{x}}(s) \le M\delta^s$, where $\delta \in (0,1)$ and $M > 0$ are some constants.*

For p-dimensional random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and $\forall j_1, \dots, j_t \in \{1, \dots, p\}$, we write the central moment as

$$\Pi_{j_1,\dots,j_t} = \mathrm{E}[(x_{j_1} - \mu_{j_1}) \dots (x_{j_t} - \mu_{j_t})]. \tag{3.6}$$

Condition 3.2.2* (Alternative dependence assumption to Condition 3.2.2). *Assume that under H_0, for any $j_1, j_2, j_3 \in \{1, \dots, p\}$, $\Pi_{j_1,j_2,j_3} = 0$; for any $j_1, j_2, j_3, j_4 \in \{1, \dots, p\}$, $\Pi_{j_1,j_2,j_3,j_4} = \kappa_1(\sigma_{j_1,j_2}\sigma_{j_3,j_4} + \sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3})$ for some constant $\kappa_1 < \infty$; and for $t = 6, 8$, and any $j_1, \cdots, j_t \in \{1, \dots, p\}$, $\Pi_{j_1,\cdots,j_t} = 0$ when at least one of these indexes appears odd times in $\{j_1, \cdots, j_t\}$.*

Condition 3.2.1 assumes that the eighth marginal moments of $\mathbf{x}$ are uniformly bounded from above and the second moments are uniformly bounded from below, which are true for most light-tailed distributions. Condition 3.2.2 assumes weak dependence among different x_j's under H_0, since the uncorrelatedness of x_j's under H_0 may not imply the independence of them, especially when x_j's are non-Gaussian. Under H_0, Condition 3.2.2 automatically holds when $\mathbf{x}$ is Gaussian or m-dependent. The mixing-type weak dependence is similarly considered in previous works (e.g., Bickel and Levina, 2008; Chen et al., 2019a; Xu et al., 2016) and also commonly assumed in time series and spatial statistics (Gaetan and Guyon, 2010; Pham and Tran, 1985). Moreover, the variables in our motivating genome-wide association studies have a local dependence structure, with their associations often decreasing to zero as the corresponding physical distances on a chromosome increase. We note that it suffices to have Condition 3.2.2 hold up to a permutation of the variables.

Alternatively, we can substitute Condition 3.2.2 with Condition 3.2.2*. Condition 3.2.2* specifies some higher order moments of $\mathbf{x}$ and is satisfied when $\mathbf{x}$ follows an elliptical distribution with finite eighth moments and covariance $\mathbf{\Sigma}$ (see Anderson, 2003; Frahm, 2004; Muirhead, 2009; Paindaveine and Van Bever, 2014). Conditions 3.2.2* and 3.2.2 become equivalent when $\mathbf{x}$ follows a multivariate Gaussian distribution. The fourth moment condition is also assumed in other high-dimensional research (Cai et al., 2013). In this work, the eighth moment condition is needed to establish the asymptotic joint distribution of different U-statistics.

The following theorem specifies the asymptotic variances of the finite order U-statistics and their joint limiting distribution. Since the U-statistics are degenerate under H_0, an analysis different from the asymptotic theory on non-degenerate U-statistics (e.g., Hoeffding, 1948) is needed in the proof.

Theorem 3.2.1. *Under H_0 in (3.2) and Conditions 3.2.1 and 3.2.2 (or 3.2.2*), for $\mathcal{U}(a)$'s defined in (3.3) and any distinct finite (and positive) integers $\{a_1, \ldots, a_m\}$, as*

$n, p \to \infty$,

$$\Big[\frac{\mathcal{U}(a_1)}{\sigma(a_1)}, \ldots, \frac{\mathcal{U}(a_m)}{\sigma(a_m)}\Big]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m), \tag{3.7}$$

where

$$\sigma^2(a) := \mathrm{var}[\mathcal{U}(a)] \simeq \frac{a!}{P^n_a} \sum_{1 \le j_1 \ne j_2 \le p;\, 1 \le j_3 \ne j_4 \le p} (\Pi_{j_1,j_2,j_3,j_4})^a, \tag{3.8}$$

with Π_{j_1,j_2,j_3,j_4} *defined in* (3.6)*. Note that* $\sigma^2(a) = \Theta(p^2 n^{-a})$*.*

Theorem 3.2.1 shows that after normalization, the finite-order U-statistics have a joint normal limiting distribution with an identity covariance matrix, which implies that they are asymptotically independent as $n, p \to \infty$. The nice independence property makes it easy to combine these U-statistics and apply our proposed adaptive testing later. Moreover, the conclusion holds on general asymptotic regime for $n, p \to \infty$, without any constraint on the relationship between n and p.

In the following, we further discuss the maximum-type test statistic $\mathcal{U}(\infty)$, which corresponds to the ℓ_∞-norm of the parameter vector $\mathcal{E} = \{e_l : l \in \mathcal{L}\}$, that is, $\|\mathcal{E}\|_\infty = \max_{l \in \mathcal{L}} |e_l|$. In the existing literature, there is already some corresponding established work (Jiang, 2004; Cai and Jiang, 2011) on the test statistic:

$$M_n^* := \max_{1 \le j_1 \ne j_2 \le p} |\hat{\sigma}_{j_1,j_2} / \sqrt{\hat{\sigma}_{j_1,j_1} \hat{\sigma}_{j_2,j_2}}|, \tag{3.9}$$

where $(\hat{\sigma}_{j_1,j_2})_{p \times p} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\intercal / n$ and $\bar{\mathbf{x}} = \sum_{i=1}^n \mathbf{x}_i / n$. We will take $\mathcal{U}(\infty) = M_n^*$ below. The limiting distribution of $\mathcal{U}(\infty)$ was first studied in Jiang (2004) and extended by Cai and Jiang (2011), Liu et al. (2008), and Shao and Zhou (2014). Next we restate the result in Cai and Jiang (2011), which gives the limiting distribution of (3.9) under the following condition.

Condition 3.2.3. *Consider the random vector* $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ *with mean vector*

$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$ *and covariance matrix* $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_{1,1}, \ldots, \sigma_{p,p})$*.* $(x_j - \mu_j)/\sqrt{\sigma_{j,j}}$ *are i.i.d. for* $j = 1, \ldots, p$*. Furthermore,* $\mathrm{E}e^{t_0(|x_1-\mu_1|/\sqrt{\sigma_{1,1}})^{\varsigma}} < \infty$ *for some* $0 < \varsigma \leq 2$ *and* $t_0 > 0$*.*

Theorem 3.2.2 (Cai and Jiang (2011, Theorem 2))**.** *Assume Condition 3.2.3 and* $\log p = o(n^{\beta})$*, where* $\beta = \varsigma/(4+\varsigma)$*. Then* $P(n \times \mathcal{U}(\infty)^2 + \varpi_p \leq u) \to G(u) = e^{-(1/\sqrt{8\pi})e^{-u/2}}$*, where* $\varpi_p = -4\log p + \log\log p$ *and* $G(u)$ *is an extreme value distribution of type I.*

Theorems 3.2.1 and 3.2.2 give the limiting distributions of $\mathcal{U}(a)$ of finite orders and $\mathcal{U}(\infty)$ respectively; it is of interest to examine their joint distribution. The following theorem shows that although $\mathcal{U}(\infty)$ has limiting distribution different from $\mathcal{U}(a)$, $a < \infty$, they are still asymptotically independent.

Theorem 3.2.3. *Assume that Condition 3.2.1 is satisfied, Condition 3.2.3 holds for* $\varsigma = 2$*, and* $\log p = o(n^{1/7})$*. For finite integers* $\{a_1, \ldots, a_m\}$*, under* H_0*,* $\mathcal{U}(a_1), \ldots, \mathcal{U}(a_m)$ *and* $\mathcal{U}(\infty)$ *are mutually asymptotically independent. In specific, for any* $z_1, \ldots, z_m, y \in \mathbb{R}$*, as* $n, p \to \infty$*,*

$$\Bigg|P\Big(n\mathcal{U}(\infty)^2 + \varpi_p \geq y, \frac{\mathcal{U}(a_1)}{\sigma(a_1)} \leq z_1, \ldots, \frac{\mathcal{U}(a_m)}{\sigma(a_m)} \leq z_m\Big) - P\Big(n\mathcal{U}(\infty)^2 + \varpi_p \geq y\Big) \times \prod_{r=1}^{m} P\Big(\frac{\mathcal{U}(a_r)}{\sigma(a_r)} \leq z_r\Big)\Bigg| \to 0.$$

Theorem 3.2.1 suggests that all the finite-order U-statistics are asymptotically independent with each other. Given this, Theorem 3.2.3 further shows that the maximum-type test statistic $\mathcal{U}(\infty)$ is also asymptotically mutually independent with those finite-order U-statistics. The conclusion shares similarity with some classical results on the asymptotic independence between the sum-of-squares-type and maximum-type statistics. Specifically, for random variables $w_1, \ldots, w_n$, Hsing (1995) and Ho and Hsing (1996) proved the asymptotic independence between $\sum_{i=1}^{n} w_i^2$ and

$\max_{i=1,\ldots,n}|w_i|$ for weakly dependent observations. The similar independence properties were extensively studied in literature (e.g. McCormick and Qi, 2000; Ho and McCormick, 1999; Peng and Nadarajah, 2003; James et al., 2007; Xu et al., 2016; Li and Xue, 2015). However, there are several differences between existing literature and the results in this chapter. First, we discuss a family of U-statistics $\mathcal{U}(a)$'s, which takes different a values, and $\mathcal{U}(2)$ here corresponding to the sum-of-squares-type statistic is only a special case of general $\mathcal{U}(a)$. Furthermore, we have shown not only the asymptotic independence between $\mathcal{U}(a)$ and $\mathcal{U}(\infty)$, but also the asymptotic independence among $\mathcal{U}(a)$'s of finite a values. Second, the constructed $\mathcal{U}(a)$'s are unbiased estimators, which are different from the sum-of-squares statistics usually examined in the literature. Moreover, the x's are allowed to be dependent and the theoretical development in the covariance testing involves a two-way dependence structure, which requires different proof techniques from the existing studies.

To apply hypothesis testing using the asymptotic results in Theorems 3.2.1 and 3.2.3, we need to estimate $\mathrm{var}\{\mathcal{U}(a)\}$. In particular, we propose the following moment estimator of (3.8):

$$\mathbb{V}_u(a) = \frac{2a!}{(P^n_a)^2} \sum_{1\le j_1\neq j_2\le p} \sum_{1\le i_1\neq\ldots\neq i_a\le n} \prod_{t=1}^{a} (x_{i_t,j_1}-\bar{x}_{j_1})^2 (x_{i_t,j_2}-\bar{x}_{j_2})^2. \qquad (3.10)$$

The next result establishes the statistical consistency of $\mathbb{V}_u(a)$.

Condition 3.2.4. *For the given integer a, $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^{8a}<\infty$.*

Theorem 3.2.4. *Under H_0 in (3.2), assume Conditions 3.2.1, 3.2.2 and 3.2.4 hold. Then $\mathbb{V}_u(a)/\mathrm{var}\{\mathcal{U}(a)\} \xrightarrow{P} 1$.*

Theorem 3.2.4 implies that the asymptotic results in Theorems 3.2.1 and 3.2.3 can still hold by replacing $\mathrm{var}\{\mathcal{U}(a)\}$ with its estimator $\mathbb{V}_u(a)$. Specifically, under H_0, $[\mathcal{U}(a_1)/\sqrt{\mathbb{V}_u(a_1)},\ldots,\mathcal{U}(a_m)/\sqrt{\mathbb{V}_u(a_m)}]^\intercal \xrightarrow{D} \mathcal{N}(0,I_m)$ under Conditions 3.2.1, 3.2.2

and 3.2.4. Moreover, Theorem 3.2.3 implies that $\{\mathcal{U}(a)/\sqrt{\mathbb{V}_u(a)}\}$'s are asymptotically independent of $\mathcal{U}(\infty)$.

3.2.2 Power Analysis

In this section, we analyze the asymptotic power of the U-statistics. The power of $\mathcal{U}(2)$ has been studied in the literature. In particular, Cai and Ma (2013) studied the hypothesis testing of a high-dimensional covariance matrix with $H_0 : \mathbf{\Sigma} = I_p$. The authors characterized the boundary that distinguishes the testable region from the non-testable region in terms of the Frobenius norm $\|\mathbf{\Sigma}-I_p\|_F$, and showed that the test statistic proposed by Chen et al. (2010) and Cai and Ma (2013), which corresponds to $\mathcal{U}(2)$ in this chapter, is rate optimal over their considered regime. However in practice, $\mathcal{U}(2)$ may be not powerful if the alternative covariance matrix is sparse with a small $\|\mathbf{\Sigma} - I_p\|_F$. When the alternative covariance has different sparsity levels, it is of interest to further examine which $\mathcal{U}(a)$ achieves the best power performance among the constructed family of U-statistics.

To study the test power, we establish the limiting distributions of $\mathcal{U}(a)$'s under the alternative hypothesis $H_A : \mathbf{\Sigma} = \mathbf{\Sigma}_A$, where the alternative covariance matrix $\mathbf{\Sigma}_A = (\sigma_{j_1,j_2})_{p\times p}$ is specified in the following Condition 3.2.5. Define $J_A = \{(j_1, j_2) : \sigma_{j_1,j_2} \neq 0, 1 \leq j_1 \neq j_2 \leq p\}$, which indicates the nonzero off-diagonal entries in $\mathbf{\Sigma}_A$. The cardinality of J_A, denoted by $|J_A|$, then represents the sparsity level of $\mathbf{\Sigma}_A$.

Condition 3.2.5. *Assume* $|J_A| = o(p^2)$ *and for* $(j_1, j_2) \in J_A$, $|\sigma_{j_1,j_2}| = \Theta(\rho)$, *where* $\rho = \sum_{(j_1,j_2)\in J_A} |\sigma_{j_1,j_2}|/|J_A|$.

Here ρ represents the average signal strength of $\mathbf{\Sigma}_A$. In our following power comparison of two U-statistics $\mathcal{U}(a)$ and $\mathcal{U}(b)$, we say $\mathcal{U}(a)$ is "better" than $\mathcal{U}(b)$, if, $\mathcal{U}(a)$ can detect a smaller average signal strength ρ than $\mathcal{U}(b)$ can detect with the same testing power; please see the definition specified by Criterion III.1 below. Condition 3.2.5 specifies a general family of "local" alternatives, which include banded covariance matrices, block covariance matrices, and sparse covariance matrices whose nonzero entries are randomly located. Moreover, we assume the following Condition 3.2.6 that can be viewed as an extension of Condition 3.2.2* to alternative settings.

Condition 3.2.6. *For $t \leq 8$, we assume that there exists a constant $\tilde{\kappa}_t$ such that $\Pi_{j_1,\ldots,j_t} = \tilde{\kappa}_t \mathrm{E}(\prod_{k=1}^{t} z_{j_k})$, where $1 \leq j_1, \ldots, j_t \leq p$ and $(z_1, \ldots, z_p)^\intercal \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_A)$.*

Similarly to Condition 3.2.2*, Condition 3.2.6 is satisfied when $\mathbf{x}$ follows an elliptical distribution with certain moment conditions (see Frahm, 2004; Maruyama and Seo, 2003).

Theorem 3.2.5. *Suppose Conditions 3.2.1, 3.2.5, and 3.2.6 hold. For $\mathcal{U}(a)$ in (3.3) and finite integers $\{a_1, \ldots, a_m\}$, if $\rho = O(|J_A|^{-1/a_t} p^{1/a_t} n^{-1/2})$ for $t = 1, \ldots, m$, then as $n, p \to \infty$,*

$$\Big[\frac{\mathcal{U}(a_1) - \mathrm{E}[\mathcal{U}(a_1)]}{\sigma(a_1)}, \ldots, \frac{\mathcal{U}(a_m) - \mathrm{E}[\mathcal{U}(a_m)]}{\sigma(a_m)}\Big]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m),$$

where for $a \in \{a_1, \ldots, a_m\}$, $\mathrm{E}[\mathcal{U}(a)] = \sum_{(j_1,j_2) \in J_A} \sigma^a_{j_1,j_2}$ and $\sigma^2(a) = \mathrm{var}[\mathcal{U}(a)] \simeq 2a!\kappa_1^a n^{-a} \sum_{1 \leq j_1 \neq j_2 \leq p} \sigma^a_{j_1,j_1}\sigma^a_{j_2,j_2}$, which is of order $\Theta(p^2 n^{-a})$.

Theorem 3.2.5 shows that for a single U-statistic $\mathcal{U}(a)$ of finite order a,

$$P\left(\frac{\mathcal{U}(a)}{\sqrt{\mathrm{var}[\mathcal{U}(a)]}} > z_{1-\alpha}\right) \to 1 - \Phi\left(z_{1-\alpha} - \frac{\mathrm{E}[\mathcal{U}(a)]}{\sqrt{\mathrm{var}[\mathcal{U}(a)]}}\right), \tag{3.11}$$

where $z_{1-\alpha}$ is the upper α quantile of $\mathcal{N}(0,1)$ and $\Phi(\cdot)$ is the cumulative distribution function of $\mathcal{N}(0,1)$. By Theorem 3.2.5, the asymptotic power of $\mathcal{U}(a)$ of the one-sided test depends on

$$\frac{\mathrm{E}[\mathcal{U}(a)]}{\sqrt{\mathrm{var}[\mathcal{U}(a)]}} \simeq \frac{\sum_{(j_1,j_2) \in J_A} \sigma^a_{j_1,j_2}}{\{2a!\kappa_1^a n^{-a} \sum_{1 \leq j_1 \neq j_2 \leq p} (\sigma_{j_1,j_1}\sigma_{j_2,j_2})^a\}^{1/2}}. \tag{3.12}$$

By Theorem 3.2.5, (3.12) $= \Theta(|J_A| \rho^a p^{-1} n^{a/2})$. It follows that when $\mathrm{E}[\mathcal{U}(a)]$ is of the same order of $\sqrt{\mathrm{var}[\mathcal{U}(a)]}$, i.e., $\mathrm{E}[\mathcal{U}(a)] = O(1)\sqrt{\mathrm{var}[\mathcal{U}(a)]}$, the constraint of ρ in Theorem 3.2.5 is satisfied.

In the following power analysis, we will first compare $\mathcal{U}(a)$'s of finite a and then compare them with $\mathcal{U}(\infty)$. As we focus on studying the relationship between the sparsity level and power, we consider an ideal case where $\sigma_{j_1,j_2} = \rho > 0$ for $(j_1, j_2) \in J_A$ and $\sigma_{j,j} = \nu^2 > 0$ for $j = 1, \ldots, p$. Then

$$(3.12) \simeq |J_A|\rho^a/(\sqrt{2a!\kappa_1^a}\nu^{2a}pn^{-a/2}). \tag{3.13}$$

We next show how the order of the "best" U-statistics changes when the sparsity level $|J_A|$ varies. To be specific of the meaning of "best", we compare the ρ values needed by different U-statistics to achieve the same asymptotic power. Particularly, we fix $\mathrm{E}[\mathcal{U}(a)]/\sqrt{\mathrm{var}[\mathcal{U}(a)]}$, i.e., (3.13) to be some constant $M/\sqrt{2}$ for different a's and the asymptotic power of each $\mathcal{U}(a)$ is $(3.11) = 1 - \Phi(z_{1-\alpha} - M/\sqrt{2})$. Then by (3.13), the ρ value such that $\mathcal{U}(a)$ attains the power above is

$$\rho_a = \sqrt{\kappa_1}(a!)^{\frac{1}{2a}}\nu^2(Mp/|J_A|)^{\frac{1}{a}}n^{-\frac{1}{2}}. \tag{3.14}$$

By the definition in (3.14), we compare the power of two U-statistics $\mathcal{U}(a)$ and $\mathcal{U}(b)$ with $a \neq b$ following the Criterion III.1 below.

Criterion III.1. We say $\mathcal{U}(a)$ is "better" than $\mathcal{U}(b)$ if $\rho_a < \rho_b$.

Given values of $n, p, |J_A|$ and M, (3.14) is a function of a. Therefore, to find the "best" $\mathcal{U}(a)$, it suffices to find the order, denoted by a_0, that gives the smallest ρ_a value in (3.14). We then have the following proposition discussing the optimality among the U-statistics of finite orders in (3.3).

Proposition 3.2.1. *Given $n, p, |J_A|$ and any constant $M \in (0, +\infty)$, we consider ρ_a in* (3.14) *as a function of integer a, then* *(i) when $|J_A| \geq Mp$, the minimum of ρ_a is achieved at $a_0 = 1$;*

(ii) when $|J_A| < Mp$, the minimum of ρ_a is achieved at some a_0, which increases as $Mp/|J_A|$ increases.

By Proposition 3.2.1, the order a_0 that attains the smallest value of ρ_a depends on the value of $Mp/|J_A|$ and does not have a closed form solution. We use numerical plots to demonstrate the relationship between a_0 and the sparsity level. Particularly, let $|J_A| = p^{2(1-\beta)}$, where $\beta \in (0, 1)$ denotes the sparsity level. To have a better visualization, we use $g(a) = \log(\rho_a n^{1/2} \kappa_1^{-1/2} \nu^{-2}) = (1/2a) \log a! + a^{-1} \log(Mp^{2\beta-1})$ instead of ρ_a. We plot $g(a)$ curves in Figure III.1 for each $\beta \in \{0.1, \ldots, 0.9\}$ with $M = 4$ and $p \in \{100, 10000\}$. Other values of M and p are also taken, which give similar patterns to Figure III.1 and are not presented.

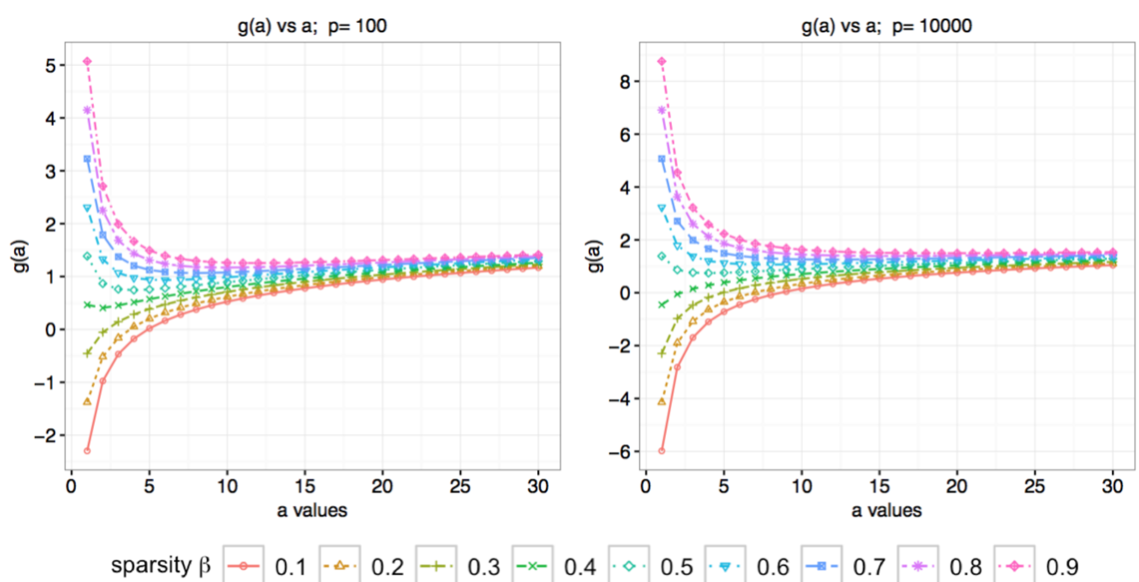


Figure III.1: $g(a)$ versus a with different sparsity level β for $p = 100$ and 10000.

Figure III.1 shows that the a_0 such that $g(a)$ attains the smallest value increases when the sparsity level β increases. In particular, when the sparsity level $\beta \leq 0.3$, that is, when $|J_A|$ is "very" large and then $\mathbf{\Sigma}_A$ is "very" dense, $g(a)$ has the smallest value at $a_0 = 1$. This is consistent with the conclusion in Proposition 3.2.1 (i). When the sparsity level β is between 0.4 and 0.5, we note that $a_0 = 2$ achieves the minimum of $g(a)$. This shows that when $|J_A|$ is "moderately" large and $\mathbf{\Sigma}_A$ is "moderately" dense, $\mathcal{U}(2)$ is more powerful than $\mathcal{U}(1)$. When the sparsity level $\beta > 0.5$, we find that $a_0 > 2$. This implies that when $|J_A|$ becomes smaller and $\mathbf{\Sigma}_A$ becomes sparser, U-statistics of higher orders are more powerful. Additionally, we note that a_0 increases slowly as β increases, which verifies Proposition 3.2.1 (ii). Moreover, the curves converge as a increases and the differences of $g(a)$ for large a values ($a \geq 6$) are small. This implies that when selecting the range of considered orders of U-statistics, it suffices to select an upper bound with $a = 6$ or 8, which gives better or similar ρ_a values to those larger a's.

In summary, when $|J_A|$ is large, i.e., $\mathbf{\Sigma}_A$ is dense, a small a tends to obtain a smaller lower bound in terms of ρ. But when $|J_A|$ decreases, i.e., $\mathbf{\Sigma}_A$ becomes sparse, a U-statistic of large finite order (or the maximum-type U-statistic as shown next) tends to obtain a smaller lower bound in ρ. This observation is consistent with the existing literature (Chen et al., 2010; Cai and Jiang, 2011; Cai and Ma, 2013; Cai, 2017).

Next, we proceed to examine the power of the maximum-type test statistic $\mathcal{U}(\infty)$, and compare it with the U-statistics $\mathcal{U}(a)$ of finite a defined in (3.3). By Cai and Jiang (2011), the rejection region for $\mathcal{U}(\infty)$ with significance level α is

$$|\mathcal{U}(\infty)| \geq t_p := n^{-1/2}\sqrt{4\log p - \log\log p - \log(8\pi) - 2\log\log(1-\alpha)^{-1}}.$$

Note $t_p \simeq 2\sqrt{\log p/n}$ and under alternative, the power for $\mathcal{U}(\infty)$ is

$$P(|\mathcal{U}(\infty)| \geq t_p). \tag{3.15}$$

As discussed, we consider the alternatives satisfying Conditions 3.2.2* and 3.2.5, $\sigma_{j_1,j_2} = \rho > 0$ for $(j_1, j_2) \in J_A$, and $\sigma_{j,j} = \nu^2$ for $j = 1, \ldots, p$. For simplicity, we assume $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$ and ν^2 are given, and focus on the simplified

$$\mathcal{U}(\infty) = \max_{1 \le j_1 < j_2 \le p} \left| \nu^{-2} n^{-1} \sum\nolimits_{i=1}^{n} (x_{i,j_1} - \mu_{j_1})(x_{i,j_2} - \mu_{j_2}) \right|. \tag{3.16}$$

We show in the following proposition when the power of $\mathcal{U}(\infty)$ asymptotically converges to 1 or is strictly smaller than 1 under alternative.

Proposition 3.2.2. *Under the considered alternative $\boldsymbol{\Sigma}_A$ above, suppose* $\max_{j=1,\dots,p} \mathrm{E}e^{t_0|x_j - \mu_j|^{\varsigma}} < \infty$ *for some* $0 < \varsigma \le 2$ *and* $t_0 > 0$*, and* $\log p = o(n^{\beta})$ *with* $\beta = \varsigma/(4+\varsigma)$*. Then for* (3.16)*, when* $n, p \to \infty$*,*

(i) there exists a constant $c_1 > 2$ *such that if* $\rho \ge c_1\sqrt{\log p/n}$*,* (3.15) $\to 1$*;*

(ii) there exists another constant $0 < c_2 < 2$ *such that when* $\rho \le c_2\sqrt{\log p/n}$*, Condition 3.2.2* holds for* $\kappa_1 \le 1$ *and* $|J_A| = o(1)p^{\frac{2(1-c_2/2)^2}{\kappa_1+m}}(\log p)^{\frac{1}{2}-\frac{1}{2(\kappa_1+m)}}$ *for some* $m > 0$*, we have* (3.15) $\le \log(1-\alpha)^{-1}$.

Recall that Proposition 3.2.1 shows that there exists a finite integer a_0, such that ρ_{a_0} is the minimum of (3.14), and ρ_{a_0} is a lower bound of ρ value for the finite-order U-statistics to achieve the given asymptotic power. With Propositions 3.2.1 and 3.2.2, we next compare the finite-order U-statistics defined in (3.3) with the maximum-type test statistic $\mathcal{U}(\infty)$.

Proposition 3.2.3. *Under the conditions of Theorem 3.2.5 and Proposition 3.2.2, for any finite integer a, there exist constants c_1 and c_2 such that when p is sufficiently large,*

(i) For any M, when $|J_A| < c_1^{-a}(a!)^{\frac{1}{2}}\kappa_1^{\frac{a}{2}}(\log p)^{-\frac{a}{2}}Mp$*,* $\mathcal{U}(\infty)$ *has higher asymptotic power than* $\mathcal{U}(a)$*.*

(ii) When M is big enough and $|J_A| > c_2^{-a}(a!)^{\frac{1}{2}}\kappa_1^{\frac{a}{2}}(\log p)^{-\frac{a}{2}}Mp$*,* $\mathcal{U}(a)$ *has higher asymptotic power than* $\mathcal{U}(\infty)$*.*

From Proposition 3.2.1, we know when $Mp/|J_A| = O(1)$, there exists a finite a_0 such that $\mathcal{U}(a_0)$ is the "best" among all the finite-order U-statistics; in this case, Proposition 3.2.3 (ii) further indicates that $\mathcal{U}(a_0)$ has higher asymptotic power than $\mathcal{U}(\infty)$. Specifically, if $Mp/|J_A| < 1$, $a_0 = 1$, then $\mathcal{U}(1)$ is the "best" and its lowest detectable order of ρ is $\Theta(p|J_A|^{-1}n^{-1/2})$. More interestingly, when $\mathbf{\Sigma}_A$ is moderately dense or moderately sparse with $Mp/|J_A| > 1$ and bounded, some U-statistic of finite order $a_0 > 1$ would become the "best". By Figure III.1, the value of a_0 increases as $\mathbf{\Sigma}_A$ becomes denser. On the other hand, when $\mathbf{\Sigma}_A$ is "very" sparse with $|J_A| < c_1^{-a_0}(a_0!)^{\frac{1}{2}}\kappa_1^{\frac{a_0}{2}}(\log p)^{-\frac{a_0}{2}}Mp$, $\mathcal{U}(\infty)$ is the "best" and its lowest detectable order of ρ is $\Theta(\sqrt{\log p/n})$.

Remark III.3. *The analysis above focuses on the ideal case where the nonzero off-diagonal entries of $\mathbf{\Sigma}_A$ are the same for illustration. When these entries of $\mathbf{\Sigma}_A$ are different, similar analysis still applies by Theorem 3.2.5 for general covariance matrices. Specifically, the asymptotic power of $\mathcal{U}(a)$ depends on the mean variance ratio (3.12) and $\rho_a = \sqrt{\kappa_1}n^{-1/2}(a!)^{1/2a} \times (M\sum_{j=1}^{p}\sigma_{j,j}^a/\sum_{1\leq j_1,j_2\leq p}\sigma_{j_1,j_2}^a)^{1/a}$. We can then obtain conclusions similar to Propositions 3.2.1–3.2.3. One interesting case is when $\mathbf{\Sigma}_A$ contains both positive and negative entries; the same analysis applies for even-order U-statistics, since σ_{j_1,j_2}^a's are all non-negative for even a. On the other hand, the odd-order U-statistics would have low power, since $\sum_{1\leq j_1\neq j_2\leq p}\sigma_{j_1,j_2}^a$ could be small due to the cancellation of positive and negative σ_{j_1,j_2}^a's. We have conducted simulations when the nonzero σ_{j_1,j_2}'s are different in Section 3.2.4, and the results exhibit consistent patterns as expected.*

3.2.3 Application to Adaptive Testing & Computation

Adaptive Testing The power analysis in Section 3.2.2 shows that when the sparsity level of the alternative changes, the test statistic that achieves the highest power could vary. However, since the truth is often unknown in practice, it is unclear which test statistic should be chosen. Therefore, we develop an adaptive testing procedure by combining the information from U-statistics of different orders, which would yield high power against various alternatives.

In particular, we propose to combine the U-statistics through their p-values, which is widely used in literature (Mosteller and Fisher, 1948; Pan et al., 2014; Yu et al., 2009). One popular method is the minimum combination, whose idea is to take the minimum p-value to approximate the maximum power (Pan et al., 2014; Yu et al., 2009; Xu et al., 2016). Specifically, let Γ be a candidate set of the orders of U-statistics, which contains both finite values and ∞. We compute p-values p_a's of the U-statistics $\mathcal{U}(a)$'s satisfying $a \in \Gamma$. The minimum combination takes the statistic $T_{\text{adpUmin}} = \min\{p_a : a \in \Gamma\}$ and has the asymptotic p-value $p_{\text{adpUmin}} = 1 - (1 - T_{\text{adpUmin}})^{|\Gamma|}$, where $|\Gamma|$ denotes the size of the candidate set Γ. We reject H_0 if $p_{\text{adpUmin}} < \alpha$. Under H_0, p_a's are asymptotically independent and uniformly distributed by the theoretical results in Section 3.2.1. The type I error is asymptotically controlled as $P(p_{\text{adpUmin}} < \alpha) = P(\min_{a\in\Gamma} p_a < p^*_\alpha) \to \alpha$, where $p^*_\alpha = 1 - (1-\alpha)^{1/|\Gamma|}$. Since $P(\min_{a\in\Gamma} p_a < p^*_\alpha) \geq P(p_a < p^*_\alpha)$, the power of the adaptive test goes to 1 if there exists $a \in \Gamma$ such that the power of $\mathcal{U}(a)$ goes to 1. We note that the power of the adaptive test is not necessarily higher than that of all the U-statistics. This is because the power of $\mathcal{U}(a)$ is $P(p_a < \alpha)$, and is different from $P(p_a < p^*_\alpha)$ since $p^*_\alpha < \alpha$ when $|\Gamma| > 1$. Based on our extensive simulations, we find that the adaptive test is usually close to or even higher than the maximum power of the U-statistics.

Remark III.4. *Fisher's method (Mosteller and Fisher, 1948) is another popular method for combining independent p-values. The test statistic* $T_{\text{adpUf}} = -2\sum_{k=1}^{|\Gamma|} \log p_k$ *converges to* $\chi^2_{2|\Gamma|}$ *under* H_0*. By our simulations, the minimum combination and Fisher's method are often comparable, while Fisher's method has higher power under several cases. Moreover, we can also use other methods to combine the p-values, such as higher criticism (Donoho and Jin, 2004, 2015). We leave the study of how to* *efficiently combine the p-values for future research.*

We select the candidate set Γ by the power analysis in Section 3.2.2. We would recommend including $\{1, 2, \ldots, 6, \infty\}$, which can be powerful against a wide spectrum of alternatives. In particular, by Propositions 3.2.1 and 3.2.3, we include $a = 1, 2$ that are powerful against dense signals; $a = \infty$ that is powerful against sparse signals; and also $a = \{3, \ldots, 6\}$ for the moderately dense and moderately sparse signals. By Figure III.1, it generally suffices to choose finite a up to 6–8, which often give similar/better performance to/than larger a values. The simulations in Section 3.2.4 confirm the good performance of this choice of Γ; and the proposed adaptive test appears to well approximate the "best" performance even when Γ may not always contain the unknown "optimal" U-statistics.

Computation Next we discuss the computation in the adaptive testing. A direct calculation following the form of $\mathcal{U}(a)$ in (3.3) and $\mathbb{V}(a)$ in (3.10) would be computationally expensive for large a with a cost of $O(p^2 n^{2a})$. To address this issue, we introduce a method that can reduce the cost.

We first consider a simplified setting when $\mathrm{E}(x_{i,j}) = 0$ to illustrate the idea. As discussed in Remark III.2, we examine $\tilde{\mathcal{U}}(a)$ defined in (3.5). Let $\mathcal{L} = \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p\}$ denote the set of index tuples, and for each index tuple $l = (j_1, j_2) \in \mathcal{L}$, define $s_{i,l} = x_{i,j_1} x_{i,j_2}$. Note that $\tilde{\mathcal{U}}(a) = (P_a^n)^{-1} \sum_{l \in \mathcal{L}} \mathcal{U}_l(a)$, where $\mathcal{U}_l(a) = \sum_{1 \leq i_1 \neq \cdots \neq i_a \leq n} \prod_{k=1}^{a} s_{i_k,l}$. Calculating $\mathcal{U}_l(a)$ directly is of order $O(n^a)$. We then focus on reducing the computational cost of $\mathcal{U}_l(a)$. For $l \in \mathcal{L}$ and finite integers $t_1, \ldots, t_k$, define

$$V_l^{(t_1,\ldots,t_k)} = \prod_{r=1}^{k} \Big(\sum_{i=1}^{n} s_{i,l}^{t_r} \Big), \quad U_l^{(t_1,\cdots,t_k)} = \sum_{1 \leq i_1 \neq \ldots \neq i_k \leq n} \prod_{r=1}^{k} s_{i_1,l}^{t_r}. \tag{3.17}$$

We can see that $\mathcal{U}_l(a) = U_l^{\mathbf{1}_a}$ with $\mathbf{1}_a$ being an a-dimensional vector of all ones, and

$U_l^{(a)} = V_l^{(a)}$ for any finite integer a. To reduce the computational cost of $\mathcal{U}_l(a)$, the main idea is to obtain $U_l^{\mathbf{1}_a}$ from $V_l^{(t_1,\ldots,t_k)}$, whose computational cost is $O(n)$. In particular, $\mathcal{U}_l(a)$ can be attained iteratively from $V_l^{(t_1,\ldots,t_k)}$ based on the following equation

$$U_l^{(k,\mathbf{1}_{r-k})} = V_l^{(k)} \times U_l^{\mathbf{1}_{r-k}} - (r-k) \times U_l^{(k+1,\mathbf{1}_{r-k-1})}, \tag{3.18}$$

which follows from the definitions. Algorithm III.1 below summarizes the steps.

Algorithm III.1: Iterative Computation Implementation

Data: $s_{i,l}$ $(1 \le i \le n,\ l \in \mathcal{L})$.
Result: $\tilde{\mathcal{U}}(a)$.
for $l \in \mathcal{L}$ **do**
  Compute and store $V_l^{(k)} = U_l^{(k)} = \sum_{i=1}^n s_{i,l}^k$, $(k = 1, \cdots, a)$ during the algorithm;
  $U_l^{\mathbf{1}_1} = V_l^{(1)}$, $U_l^{\mathbf{1}_2} = U_l^{\mathbf{1}_1} V_l^{(1)} - U_l^{(2)}$;
  while $3 \le r \le a$ **do**
    $T_l = U_l^{(r)}$
    for $k \leftarrow r-1$ **to** 1 **do**
      $T_l = V_l^{(k)} \times U_l^{\mathbf{1}_{r-k}} - (r-k) \times T_l$
    end
    $U_l^{\mathbf{1}_r} = T_l$
  end
end
$\tilde{\mathcal{U}}(a) = (P_a^n)^{-1} \sum_{l \in \mathcal{L}} U_l^{\mathbf{1}_a}$

We illustrate the idea of the algorithm by some examples. By definition, $U_l^{(1)} = V_l^{(1)}$, which can be computed with cost $O(n)$. Next consider in (3.18), if $r = 2$ and $k = 1$, then $U_l^{(1,1)} = V_l^{(1)} \times U_l^{(1)} - (2-1) \times U_l^{(2)} = V_l^{(1)} \times V_l^{(1)} - V_l^{(2)}$, which yields $U_l^{\mathbf{1}_2}$ with cost $O(n)$. For $U_l^{\mathbf{1}_3}$, we first take $r = 3$ and $k = 2$ in (3.18), then with cost $O(n)$, we have $U_l^{(2,1)} = V_l^{(2)} \times U_l^{(1)} - U_l^{(3)} = V_l^{(2)} \times V_l^{(1)} - V_l^{(3)}$, as $V_l^{(k)} = U_l^{(k)}$ by the definition. Given $U_l^{\mathbf{1}_2}$ and $U_l^{(2,1)}$, we obtain $U_l^{(1,\mathbf{1}_2)} = V_l^{(1)} \times U_l^{\mathbf{1}_2} - 2 \times U_l^{(2,\mathbf{1}_1)}$. Thus $U_l^{\mathbf{1}_3}$ is also computed with cost $O(n)$. Iteratively, for any finite integer a, we can obtain $U_l^{\mathbf{1}_a}$ from $V_l^{(t_1,\ldots,t_k)}$ whose computational cost is $O(n)$. More closed form formulae representing $U_l^{\mathbf{1}_a}$ by $V_l^{(t_1,\ldots,t_k)}$ are given in Section B.6.1.

Algorithm III.1 reduces the computational cost of $\tilde{\mathcal{U}}(a)$ from $O(p^2n^a)$ to $O(p^2n)$. Its idea is general and can be extended to compute other different U-statistics by changing the input $s_{i,l}$. In particular, the variance estimator $\mathbb{V}(a)$ can be computed with cost $O(p^2n)$ by specifying $s_{i,l} = (x_{i,j_1} - \bar{x}_{j_1})^2(x_{i,j_2} - \bar{x}_{j_2})^2$, for each $l \in \mathcal{L} = \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p\}$. Then $\mathbb{V}(a) = 2a!(P_a^n)^{-2} \sum_{l\in\mathcal{L}} \sum_{1\leq i_1\neq\ldots\neq i_a\leq n} \prod_{k=1}^a s_{i_k,l}$ and the Algorithm III.1 can be applied. Moreover, when $\mathrm{E}(x_{i,j})$ is unknown, $\mathcal{U}(a)$ can still be computed with cost $O(p^2n)$ using the iterative method similar to Algorithm III.1. The details are provided in Section B.6.2.

3.2.4 Simulation Studies

We conduct simulation studies to evaluate the performance of the proposed adaptive testing procedures, and investigate the relationship between the power and sparsity levels. For one-sample covariance testing discussed in Section 3.2, we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \ldots, n$, and consider the following five simulation settings.

Setting 1: $\mathbf{x}_i$ has p i.i.d. entries of $\mathcal{N}(0, 1)$ and Gamma$(2, 0.5)$ respectively. Under each case, we take $n = 100$ and $p \in \{50, 100, 200, 400, 600, 800, 1000\}$ to verify the theoretical results under H_0 and the validity of the adaptive test across different n and p combinations.

For the following settings 2–5, we generate $\mathbf{x}_i$ from multivariate Gaussian distributions with mean zero and different covariance matrices $\mathbf{\Sigma}_A$'s.

Setting 2: $\mathbf{\Sigma}_A = (1-\rho)I_p + \rho\mathbf{1}_{p,k_0}\mathbf{1}_{p,k_0}^\intercal$, where $\mathbf{1}_{p,k_0}$ is a p-dimensional vector with the first k_0 elements one and the rest zero. We take $(n, p) \in \{(100, 300), (100, 600), (100, 1000)\}$, and study the power with respect to different signal sizes ρ and sparsity levels k_0.

Setting 3: The diagonal elements of $\mathbf{\Sigma}_A$ are all one and $|J_A|$ number of off-diagonal elements are ρ with random positions. We take $(n, p) \in \{(100, 600), (100, 1000)\}$ and let the signal size ρ and sparsity level $|J_A|$ vary to examine how the power changes accordingly.

Setting 4: The diagonal elements of $\mathbf{\Sigma}_A$ are all one and $|J_A|$ number of off-diagonal elements are uniformly generated from $(0, 2\rho)$ with random positions. We take $(n, p) = (100, 1000)$ and similarly let the signal size ρ and sparsity level $|J_A|$ vary to examine how the power changes accordingly.

Setting 5: We consider the multivariate models in Chen et al. (2010). Specifically, for each $i = 1, \dots, n$, $\mathbf{x}_i = \Xi \mathbf{z}_i + \boldsymbol{\mu}$, where Ξ is a matrix of dimension $p \times m$, and $\mathbf{z}_i$'s are i.i.d. Gaussian or Gamma random vectors. Under null hypothesis, $m = p$, $\Xi = I_p$ $\boldsymbol{\mu} = 2\mathbf{1}_p$; under alternative hypothesis, $m = p + 1$, $\Xi = (\sqrt{1-\rho} I_p, \sqrt{2\rho}\mathbf{1}_p)$, $\boldsymbol{\mu} = 2(\sqrt{1-\rho} + \sqrt{2\rho})\mathbf{1}_p$. We also take the n and p combination in Chen et al. (2010) with $(n, p) \in \{(40, 159), (40, 331), (80, 159), (80, 331), (80, 642)\}$.

We compare several methods in the literature, including both maximum-type and sum-of-squares-type tests. In particular, the maximum-type test statistic in Jiang (2004) is taken as $\mathcal{U}(\infty)$ in this framework. Since the convergence in Jiang (2004) is known to be slow, we use permutation to approximate the distribution in the simulations. In addition, we consider some sum-of-squares-type methods. Specifically, we examine the identity and sphericity tests in Chen et al. (2010), which are denoted as "Equal" and "Spher", respectively. We also compare the methods in Ledoit and Wolf (2002) and Schott (2007), which are referred to as "LW" and "Schott", respectively.

To illustrate, Figure III.2 summarizes the numerical results for the setting 3 when $n = 100$ and $p = 1000$. All the results are based on 1000 simulations at the 5% nominal significance level. In Figure III.2, we present the power of single U-statistics with orders in $\{1, \dots, 6, \infty\}$. "adpUmin" and "adpUf" represent the results of the adaptive testing procedure using the minimum combination and Fisher's method in Section 3.2.2 respectively. The simulation results show that the type I error rates of the U-statistics and adaptive test are well controlled under H_0. In addition, Figure III.2 exhibits several patterns that are consistent with the power analysis in Section 3.2.2. First, it shows that among the U-statistics, when $|J_A|$ is very small, $\mathcal{U}(\infty)$ performs best; and when $|J_A|$ increases, the performances of some U-statistics of finite orders catch up. For instance, when $|J_A| = 100$, $\mathcal{U}(6)$ and $\mathcal{U}(\infty)$ are similar and are better than the other U-statistics; when $|J_A| = 400$, $\mathcal{U}(4)$ and $\mathcal{U}(5)$ are similar and better than the other U-statistics. When $\mathbf{\Sigma}_A$ is relatively dense, $\mathcal{U}(2)$ and $\mathcal{U}(1)$ become more powerful. Particularly, when $|J_A| = 1600$, $\mathcal{U}(2)$ is powerful; when $|J_A|$ becomes larger, such as when $|J_A| = 3200$, $\mathcal{U}(1)$ is overall the most powerful. Second, Figure III.2 shows that "LW", "Schott", "Equal", "Spher" and $\mathcal{U}(2)$ perform similarly under various cases. In particular, these methods are not powerful when the alternative is sparse but becomes more powerful when the alternative gets denser. This is because they are all sum-of-squares-type statistics that target at dense alternatives. Third and importantly, the two adaptive tests "adpUmin" and "adpUf" maintain high power across different settings. Specifically, they perform better than most single U-statistics: their powers are usually close to or even higher than the best single U-statistic. Moreover, "adpUmin" and "adpUf" generally have higher power than the compared existing methods. We also note that "adpUf" overall performs better than "adpUmin" in this simulation setting. In summary, Figure III.2 demonstrates the relationship between the sparsity levels of alternatives and the power of the tests, confirming the theoretical conclusions in Section 3.2.2. Notably, the proposed adaptive testing procedure is powerful against a wide range of alternatives, and thus advantageous in practice when the true alternative is unknown.

Moreover, we provide other extensive numerical studies in Section B.7.1. The conclusions are similar to those of Figure III.2, and consistent with the theoretical results in Section 3.2.2. In particular, the results show that the empirical sizes of the tests are close to the nominal level, suggesting the good finite-sample performance

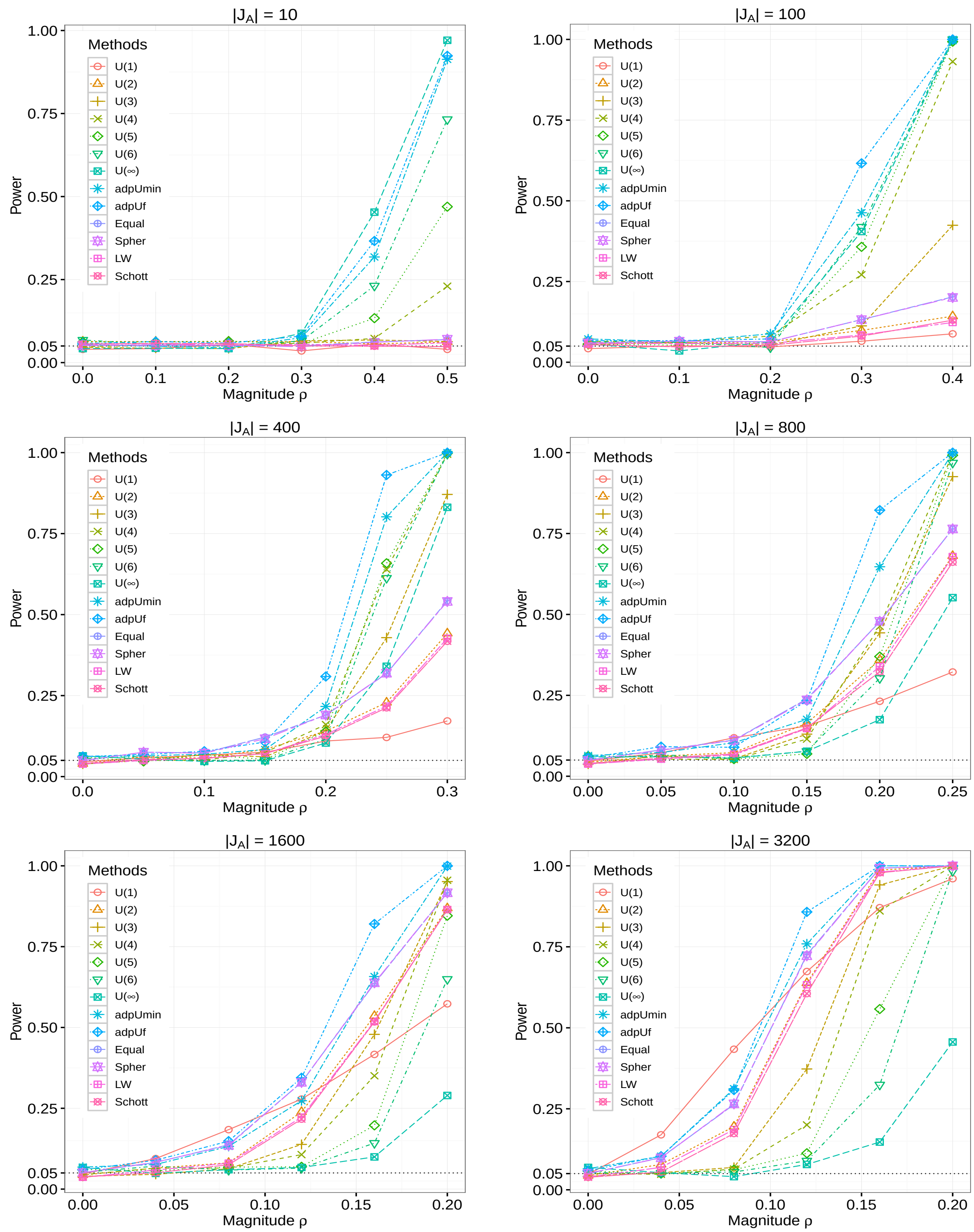


Figure III.2: Power comparison of different one-sample covariance tests under the setting 3 with $n = 100$ and $p = 1000$.

of the asymptotic approximations. Moreover, under highly dense alternatives with only non-negative entries in the covariance matrix, $\mathcal{U}(1)$ is the most powerful one among the $\mathcal{U}(a)$'s and the other tests in Ledoit and Wolf (2002), Schott (2007), and Chen et al. (2010), in agreement with the results in Propositions 3.2.1 and 3.2.3. Furthermore, the proposed adaptive testing procedures often have higher power than most single U-statistics.

3.2.5 Data Example: Alzheimer's Disease Neuroimaging Initiative

Alzheimer's disease (AD) is the most prevalent neurodegenerative disease (Prince et al., 2013) and is ranked as the sixth leading cause of death in the US (Xu et al., 2018). Every 65 seconds, someone in the US develops AD (Alzheimer's Association, 2018). To advance our understanding of AD, the Alzheimer's Disease Neuroimaging Initiative (ADNI) was started in 2004, collecting extensive genetic data for both healthy individuals and AD patients. To gain insight into the genetic mechanisms of AD, one can test a single SNP a time. However, due to a relatively small sample size of the ADNI data, scanning across all SNPs failed to identify any genome-wide significant SNP (with p-value $< 5\times10^{-8}$)(Kim et al., 2016). To date, the largest meta-analysis of more than 600,000 individuals identified 29 significant risk loci (Jansen et al., 2019) and can only explain a small proportion of AD variance. On the other hand, a group of functionally related genes as annotated in a biological pathway are often involved in the same disease susceptibility and progression (Heinig et al., 2010). Thus, pathway-based analyses, which jointly analyze a group of SNPs in a biological pathway, have become increasingly popular. We retrieve a total of 214 pathways from the KEGG database (Kanehisa et al., 2010) for the subsequent analysis.

Although pathway-based analyses with KEGG pathways are common in real studies, formally testing the correlations of the genes in a KEGG pathway has been largely untouched. Here, we apply our method and other competing methods in Chen et al. (2010) to test if all the genes in a pathway have correlated gene expression levels. Perhaps as expected, all methods reject the null hypothesis for all pathways with highly significant p-values, since the KEGG pathways are constructed to include only the genes with similar function into the same pathway (Kanehisa et al., 2010), while similar function often implies co-expression (and vice versa). To compare the performance of the different tests, for each pathway we randomly select 50 subjects and restrict our analysis to pathways of at least 50 genes, leading to 103 pathways for the following analysis. Then we perturb the data by shuffling the gene expression levels of randomly selected $100(1-\alpha)\%$ genes in a pathway before applying each test. Figure III.3 shows the performance of the tests with two significance cutoffs, where "$\mathcal{U}(2)$" represents the single $\mathcal{U}(2)$ statistic, "adpU" represents our proposed adaptive testing procedure using the minimum combination with candidate U-statistics of orders in $\{1, \ldots, 6, \infty\}$, and "Equal" and "Spher" represent the identity and sphericity tests in Chen et al. (2010) respectively. Because all pathways are highly significant with all samples, we can treat all pathways as the true positives. Due to the adaptiveness of our proposed testing procedure, "adpU" identifies more significant pathways than the competing methods across all the levels of data perturbation (mimicking the varying sparsity levels of the alternatives).

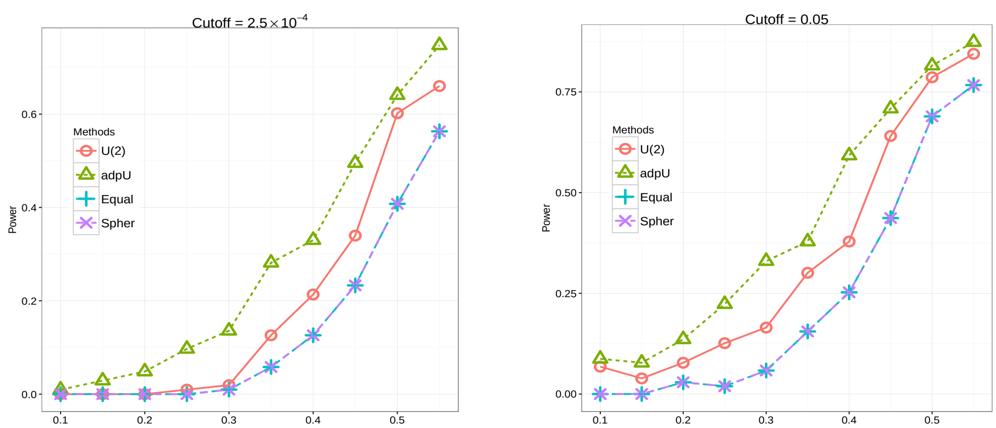


Figure III.3: Power comparison of different one-sample covariance tests with ADNI data.

3.3 One-Sample and Two-Sample Mean Tests

Testing mean vectors is widely used in many statistical analysis and applications (Anderson, 2003; Muirhead, 2009). Under high-dimensional scenarios, e.g., in genome-wide studies, dimension of the data is often much larger than the sample size, so traditional multivariate tests such as Hotelling's T^2-test either cannot be directly applied or have low power (Fan, 1996). To address this issue, several new procedures for testing high-dimensional mean vectors have been proposed (Bai and Saranadasa, 1996; Donoho and Jin, 2004; Goeman et al., 2006; Srivastava and Du, 2008; Chen and Qin, 2010; Hall and Jin, 2010; Cai et al., 2014; Chen et al., 2019a; Gregory et al., 2015; Donoho and Jin, 2015; Srivastava et al., 2016; Xu et al., 2016). However, many of the statistics only target at either sparse or dense alternatives, and suffer from loss of power for other types of alternatives. We next apply the U-statistics framework to one-sample and two-sample mean testing problems.

One-sample mean test We first discuss the one-sample mean vector testing. Assume that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are n i.i.d. copies of a p-dimensional real-valued random vector $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$, covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{j_1,j_2} : 1 \le j_1, j_2 \le p\}$. We want to conduct the global test on $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ where $\boldsymbol{\mu}_0 = (\mu_{1,0}, \ldots, \mu_{p,0})^\intercal$ is given.

Similar to previous discussion, the parameter set that we are interested in is $\mathcal{E} = \{\mu_1 - \mu_{1,0}, \ldots, \mu_p - \mu_{p,0}\}$. For each $j = 1, \ldots, p$, $\mathrm{E}(x_{i,j}) = \mu_j$, so $K_j(\mathbf{x}_i) = x_{i,j} - \mu_{j,0}$ is a kernel function, which is a simple unbiased estimator of the target. Following our construction, the U-statistic for finite a is

$$\mathcal{U}(a) = \sum_{j=1}^{p} \frac{1}{P_a^n} \sum_{1 \le i_1 \ne \cdots \ne i_a \le n} \prod_{k=1}^{a} (x_{i_k,j} - \mu_{j,0}), \tag{3.19}$$

which targets at $\|\mathcal{E}\|_a^a = \sum_{j=1}^{p} (\mu_j - \mu_{j,0})^a$, and the U-statistic corresponding to $\|\mathcal{E}\|_\infty$

is $\mathcal{U}(\infty) = \max_{1\le j\le p} \sigma_{j,j}^{-1}(\bar{x}_j - \mu_{0,j})^2$ with $\bar{x}_j = \sum_{i=1}^n x_{i,j}/n$.

Given the statistics, we have the theoretical results similar to Theorems 3.2.1–3.2.3. The following Theorems 3.3.1–3.3.2 are established under similar conditions to that of Theorems 3.2.1–3.2.3.

Condition 3.3.1.

(1) $\lim_{p\to\infty} \max_{1\le j\le p} \mathrm{E}(x_j - \mu_j)^4 < \infty$*;* $\lim_{p\to\infty} \min_{1\le j\le p} \mathrm{E}(x_j - \mu_j)^2 > 0$.

(2) $\mathbf{x}$ *is α-mixing with $\alpha_x(s) \le M\delta^s$, where $\delta \in (0,1)$ and $M > 0$ are some constants. In addition, $\sum_{j_1,j_2=1}^p \sigma_{j_1,j_2}^a = \Theta(p)$.*

Condition 3.3.1 is similar to Conditions 3.2.1 and 3.2.2 of Theorem 3.2.1. As the mean is a lower order moment function than the covariance, Condition 3.3.1 (1) is weaker than Condition 3.2.1 in that only the fourth moments are needed to be uniformly bounded instead of the eighth moments. Condition 3.3.1 (2) is a regularization condition of the structure of the covariance matrix.

Theorem 3.3.1. *Under H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, assume Condition 3.3.1. Then for any finite integers $\{a_1,\ldots,a_m\}$, as $n,p\to\infty$, $[\mathcal{U}(a_1)/\sigma(a_1),\ldots,\mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, where $\sigma^2(a) = \mathrm{var}[\mathcal{U}(a)] = \sum_{i=1}^p\sum_{j=1}^p a!\sigma_{i,j}^a/P_a^n$ with the order of $\Theta(a!pn^{-a})$.*

Condition 3.3.2.

(1) There exists constant B such that $B^{-1} \le \lambda_{\min}(\boldsymbol{\Sigma}) \le \lambda_{\max}(\boldsymbol{\Sigma}) \le B$, where $\lambda_{\min}(\boldsymbol{\Sigma})$ and $\lambda_{\max}(\boldsymbol{\Sigma})$ denote the minimum and maximum eigenvalues of the covariance matrix $\boldsymbol{\Sigma}$; and all correlations are bounded away from -1 and 1, i.e., $\max_{1\le j_1\ne j_2\le p} |\sigma_{j_1,j_2}|/(\sigma_{j_1,j_2}\sigma_{j_2,j_2})^{1/2} < 1-\eta$ for some $\eta > 0$.

(2) $\log p = o(n^{1/4})$*;* $\max_{1\le j\le p} \mathrm{E}[\exp(h(x_j-\mu_j)^2)] < \infty$, *for $h \in [-M_1, M_1]$, where $M_1 > 0$ is some constant.*

(3) $\{(x_{i,j}, i = 1,\ldots,n) : 1 \le j \le p\}$ *is α-mixing with $\alpha_x(s) \le C\delta^s$, where $\delta \in (0,1)$ and $C > 0$ is some constant, and $\sum_{j_1,j_2=1}^p \sigma_{j_1,j_2}^a = \Theta(p)$.*

In Condition 3.3.2, (1) and (2) are assumed to establish the extreme value distribution of $\mathcal{U}(\infty)$, as in Cai et al. (2014) and Xu et al. (2016). Furthermore, the mixing condition in Condition (3) is used to establish the joint independence of finite order U-statistics and $\mathcal{U}(\infty)$, following the argument in Hsing (1995).

Theorem 3.3.2. *Under H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, assume Condition 3.3.2. Then $\forall u \in \mathbb{R}$, $P(n\mathcal{U}(\infty) - \tau_p \leq u) \to \exp\{-\pi^{-1/2}\exp(-u/2)\}$, as $n, p \to \infty$, where $\tau_p = 2\log p - \log\log p$. In addition, for any finite integer a, $\{\mathcal{U}(a)/\sigma(a)\}$ and $\{n\mathcal{U}(\infty) - \tau_p\}$ are asymptotically independent.*

By Theorems 3.3.1 and 3.3.2, we obtain the asymptotic independence among the U-statistics and the corresponding limiting distributions of the U-statistics under H_0. Under the alternative hypothesis, since the power analysis of the one-sample mean testing is similar to that of the two-sample case, we delay the power analysis after presenting the asymptotic independence property of the proposed U-statistics in the two-sample mean testing problem.

Two-sample mean test Next we discuss the two-sample mean testing problem. Suppose we have two groups of p-dimensional observations $\{\mathbf{x}_i\}_{i=1}^{n_x}$ and $\{\mathbf{y}_i\}_{i=1}^{n_y}$, which are i.i.d. copies of two independent random vectors $\mathbf{x} = (x_1, \ldots, x_p)^\intercal$ and $\mathbf{y} = (y_1, \ldots, y_p)^\intercal$ respectively. Suppose $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\intercal$, $\mathrm{E}(\mathbf{y}) = \boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)^\intercal$, $\mathrm{cov}(\mathbf{x}) = \boldsymbol{\Sigma}_x$ and $\mathrm{cov}(\mathbf{y}) = \boldsymbol{\Sigma}_y$. We write $n = n_x + n_y$ and assume $n_x = \Theta(n_y)$. For easy illustration, we first consider $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_y = \boldsymbol{\Sigma} = \{\sigma_{j_1,j_2} : 1 \leq j_1, j_2 \leq p\}$. We will then discuss the case when $\boldsymbol{\Sigma}_x \neq \boldsymbol{\Sigma}_y$, where similar analysis applies.

The two-sample mean testing examines $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{\nu}$, then $\mathcal{E} = (\mu_1 - \nu_1, \ldots, \mu_p - \nu_p)^\intercal$. For $1 \leq j \leq p$, $1 \leq k \leq n_x$, $1 \leq s \leq n_y$, $K_j(\mathbf{x}_k, \mathbf{y}_s) = x_{k,j} - y_{s,j}$ is a simple unbiased estimator of $\mu_j - \nu_j$, and thus we construct $\mathcal{U}(a) =$

$\sum_{j=1}^{p}(P_a^{n_x}P_a^{n_y})^{-1}\sum_{\substack{1\le k_1\neq\ldots\neq k_a\le n_x\\ 1\le s_1\neq\ldots\neq s_a\le n_y}}\prod_{t=1}^{a}(x_{k_t,j}-y_{s_t,j})$, which is also equivalent to

$$\mathcal{U}(a)=\sum_{j=1}^{p}\sum_{c=0}^{a}\binom{a}{c}\frac{(-1)^{a-c}}{P_c^{n_x}P_{a-c}^{n_y}}\sum_{\substack{1\le k_1\neq\cdots\neq k_c\le n_x\\ 1\le s_1\neq\cdots\neq s_{a-c}\le n_y}}\prod_{t=1}^{c}x_{k_t,j}\prod_{m=1}^{a-c}y_{s_m,j}. \tag{3.20}$$

We can check that (3.20) satisfies $\mathrm{E}\{\mathcal{U}(a)\}=\sum_{j=1}^{p}(\mu_j-\nu_j)^a$, so $\mathcal{U}(a)$ is an unbiased estimator of $\|\mathcal{E}\|_a^a=\sum_{j=1}^{p}(\mu_j-\nu_j)^a$. On the other hand, for $\|\mathcal{E}\|_\infty$, following the maximum-type test statistic in Cai et al. (2014), we have

$$\mathcal{U}(\infty)=\max_{1\le j\le p}\sigma_{j,j}^{-1}(\bar{x}_j-\bar{y}_j)^2, \tag{3.21}$$

where $\bar{x}_j=\sum_{i=1}^{n_x}x_{i,j}/n_x$, $\bar{y}_j=\sum_{i=1}^{n_y}y_{i,j}/n_y$. We then obtain results similar to Theorems 3.2.1, 3.2.3 and 3.2.5 under conditions similar to those in Section 3.2.

Condition 3.3.3.

(1) There exists constant B such that $B^{-1}\le\lambda_{\min}(\mathbf{\Sigma}_x)\le\lambda_{\max}(\mathbf{\Sigma}_x)\le B$, where $\lambda_{\min}(\mathbf{\Sigma}_x)$ and $\lambda_{\max}(\mathbf{\Sigma}_x)$ denote the minimum and maximum eigenvalues of $\mathbf{\Sigma}_x$; and $\max_{1\le j_1\neq j_2\le p}|\sigma_{x,j_1,j_2}|/(\sigma_{x,j_1,j_2}\sigma_{x,j_2,j_2})^{1/2}<1-\eta$ for some $\eta>0$, i.e., all correlations are bounded away from -1 and 1. In addition, we assume the same assumptions hold for $\mathbf{\Sigma}_y$.

(2) $n,p\to\infty$, $\log p=o(1)n^{1/4}$ and $n_x/n\to\gamma\in(0,1)$. Also, $\max_{1\le j\le p}\mathrm{E}[\exp(h(x_j-\mu_j)^2)]<\infty$ and $\max_{1\le j\le p}\mathrm{E}[\exp(h(y_j-\nu_j)^2)]<\infty$, for $h\in[-M,M]$, where M is a positive constant.

(3) $\{(x_{i,j},i=1,\ldots,n):1\le j\le p\}$ and $\{(y_{i,j},i=1,\ldots,n):1\le j\le p\}$ are α-mixing with $\alpha_x(s)\le C\delta_x^s$ and $\alpha_y(s)\le C\delta_y^s$, where $\delta_x,\delta_y\in(0,1)$ and C is some constant. We also assume $\sum_{j_1,j_2=1}^{p}\{\sigma_{x,j_1,j_2}/\gamma+\sigma_{y,j_1,j_2}/(1-\gamma)\}^a=\Theta(p)$.

Condition 3.3.3 is similar to Condition 3.3.2. They are assumed to establish both the limiting distributions and asymptotic independence properties of $\mathcal{U}(a)$ and $\mathcal{U}(\infty)$

for testing two-sample mean.

Theorem 3.3.3. *Under Condition 3.3.3, $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y$ and $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$, for any finite integers $(a_1, \ldots, a_m)$, as $n, p \to \infty$, $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, where $\sigma^2(a) \simeq a! \sum_{j_1,j_2=1}^{p} (n_x + n_y)^a \sigma_{j_1,j_2}^a/(n_x n_y)^a$ is of the order $\Theta(a!pn^{-a})$.*

Theorem 3.3.4. *Under Condition 3.3.3, $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y$ and $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$, $\forall u \in \mathbb{R}$, $P(\frac{n_x n_y}{n_x+n_y}\mathcal{U}(\infty) - \tau_p \leq u) \to \exp\{-\pi^{-1/2}\exp(-u/2)\}$, as $n, p \to \infty$, where $\tau_p = 2\log p - \log\log p$. Moreover, $\{\mathcal{U}(a)/\sigma(a)\}$ of finite integer a and $\{n_x n_y \mathcal{U}(\infty)/(n_x + n_y) - \tau_p\}$ are asymptotically independent.*

Theorems 3.3.3 and 3.3.4 provide the asymptotic properties of finite-order U-statistics and $\mathcal{U}(\infty)$ under H_0. To analyze the power of $\mathcal{U}(a)$'s, we derive the asymptotic results of $\mathcal{U}(a)$'s under the alternative hypotheses. We focus on the two-sample mean testing problem, while one-sample mean testing can be obtained similarly. Specifically, we consider the alternative $\mathcal{E}_A = \{\mu_j - \nu_j = \rho > 0 \text{ for } j = 1, \ldots, k_0; \mu_j - \nu_j = 0 \text{ for } j = k_0 + 1, \cdots, p\}$. We then obtain similar conclusions to Theorem 3.2.5.

Theorem 3.3.5. *Assume Condition 3.3.3 and $k_0 = o(p)$. For any finite integers $\{a_1, \ldots, a_m\}$, if ρ in $\mathcal{E}_A$ satisfies $\rho = O(k_0^{-1/a_t} p^{1/(2a_t)} n^{-1/2})$ for $t = 1, \ldots, m$, then $[\mathcal{U}(a_1) - \mathrm{E}\{\mathcal{U}(a_1)\}]/\sigma(a_1), \ldots, [\mathcal{U}(a_m) - \mathrm{E}\{\mathcal{U}(a_m)\}]/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, as $n, p \to \infty$. Here $\mathrm{E}[\mathcal{U}(a)] = \|\mathcal{E}_A\|_a^a = k_0\rho^a$ and $\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\} \simeq V_a$, with $V_a = a! \sum_{j_1,j_2=k_0+1}^{p} (n_x + n_y)^a \sigma_{j_1,j_2}^a/(n_x n_y)^a$ of the order $\Theta(a!pn^{-a})$.*

Next we compare the power of different U-statistics under alternatives with different sparsity levels. Theorem 3.3.5 shows that under the local alternatives, the asymptotic power of $\mathcal{U}(a)$ mainly depends on $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$. Therefore by Theorem 3.3.5, given constant $M > 0$, for each $\mathcal{U}(a)$, if $\rho = M^{1/a} k_0^{-1/a} V_a^{1/(2a)}$, then $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$; that is, different $\mathcal{U}(a)$'s have the same power asymptotically. For easy illustration, we consider $\sigma_{j_1,j_2} = 1$ when $j_1 = j_2 \in \{k_0 + 1, \ldots, p\}$, and $\sigma_{j_1,j_2} = 0$ when $j_1 \neq j_2 \in \{k_0 + 1, \ldots, p\}$, then $M^{1/a} k_0^{-1/a} V_a^{1/(2a)} \simeq \rho_a$ with

$$\rho_a := a!^{\frac{1}{2a}} (M\sqrt{p}/k_0)^{\frac{1}{a}} \{(n_x + n_y)/(n_x n_y)\}^{\frac{1}{2}}. \tag{3.22}$$

Therefore, similarly to the analysis in Section 3.2.2, to find the "best" $\mathcal{U}(a)$, it suffices to find the order, denoted by a_0, that gives the minimum ρ_a in (3.22). We have the following result similar to Proposition 3.2.1.

Proposition 3.3.1. *Given any constant $M \in (0, +\infty)$ and n, p, k_0, we consider ρ_a in* (3.22) *as a function of positive integers a, then*

(i) when $k_0 \geq M\sqrt{p}$, the minimum of ρ_a is achieved at $a_0 = 1$;

(ii) when $k_0 < M\sqrt{p}$, the minimum of ρ_a is achieved at some a_0, which increases as $M\sqrt{p}/|J_D|$ increases.

Proposition 3.3.1 shows that when the sparsity level k_0 is large, i.e., $\mathcal{E}_a$ is dense, a small a tends to obtain a smaller lower bound in ρ, and vice versa. As (3.22) and (3.14) are similar, we have similar patterns to that in Figure III.1 when examining the corresponding numerical plots of ρ_a. In addition, Cai et al. (2014) shows that when $\rho = \rho_\infty := C_1\sqrt{\log p/n}$ for a large C_1, the power of $\mathcal{U}(\infty)$ converges to 1, and $\sqrt{\log p/n}$ is minimax rate optimal for sparse alternatives; see also Donoho and Jin (2015). Thus, if $\rho_\infty < \rho_{a_0}$, i.e., $k_0 < MC_1^{-a_0}\sqrt{pa_0!}/\log^{a_0/2} p$, $\mathcal{U}(\infty)$ is the "best" and its lowest detectable order of ρ is $\Theta(\sqrt{\log p/n})$. On the other hand, Proposition 3.3.1 shows that when $\mathcal{E}_A$ is dense with $k_0 > \sqrt{Mp}$, $\mathcal{U}(1)$ is the "best" and its lowest detectable order of ρ is $\Theta(\sqrt{p}k_0^{-1}n^{-1/2})$. Moreover, for some large M and C_2, when $\mathcal{E}_A$ is "moderately dense" or "moderately sparse" with $C_2\sqrt{pa_0!}/\log^{a_0/2} p < k_0 < \sqrt{Mp}$, $\mathcal{U}(a_0)$ is the "best" and its lowest detectable order of ρ is $\Theta\{(\sqrt{p}/k_0)^{\frac{1}{a_0}} n^{-1/2}\}$, which is of a smaller order than the optimal detection boundary of the sparse case $\Theta(\sqrt{\log p/n})$.

More generally, when $\mathbf{\Sigma}_x \neq \mathbf{\Sigma}_y$, similar results to Theorems 3.3.3 and 3.3.5 can be obtained. In particular, we have the following corollary.

Corollary 3.3.1. *When $\mathbf{\Sigma}_x \neq \mathbf{\Sigma}_y$, under Condition 3.3.3, Theorem 3.3.3 holds with $\sigma^2(a) \simeq a! \sum_{j_1,j_2=1}^{p} (\sigma_{x,j_1,j_2}/n_x + \sigma_{y,j_1,j_2}/n_y)^a$ and Theorem 3.3.5 holds with $V_a = a! \sum_{j_1,j_2=k_0+1}^{p} (\sigma_{x,j_1,j_2}/n_x + \sigma_{y,j_1,j_2}/n_y)^a$.*

Corollary 3.3.1 shows that the asymptotic power of finite-order U-statistics depends on $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$. By the construction of finite-order U-statistics and the proof, we obtain that $\mathrm{E}\{\mathcal{U}(a)\} = k_0 \rho^a$ and $\mathrm{var}\{\mathcal{U}(a)\} = \Theta(a! p n^{-a})$. We then know that for finite-order U-statistics, similar results to Proposition 3.3.1 still hold by examining $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$.

The above power analysis shows that the optimal U-statistic varies when the alternative hypothesis changes. To achieve high power across various alternatives, we can develop an adaptive test similar to that in Section 3.2.3. Specifically, we calculate the p-values of the U-statistics (3.19) and (3.20) following the theoretical results above and the algorithm in Section 3.2.3. By combining the p-values as discussed in Section 3.2.3, the asymptotic power of the adaptive test goes to 1 if there exists one $\mathcal{U}(a)$ whose power goes to 1.

Remark III.5. *Xu et al. (2016) has also discussed the adaptive testing of two-sample mean that is powerful against various ℓ_p-norm-like sums of $\boldsymbol{\mu} - \boldsymbol{\nu}$. But Xu et al. (2016) is under the framework of a family of von Mises V-statistics where $\mathcal{V}(a) = \sum_{j=1}^{p} (\bar{x}_j - \bar{y}_j)^a$. We note that $\mathcal{V}(a)$ is equivalent to*

$$\mathcal{V}(a) = \sum_{j=1}^{p} \sum_{c=0}^{a} (-1)^{a-c} \binom{a}{c} ({n_x}^c {n_y}^{a-c})^{-1} \sum_{\substack{1 \leq k_1, \cdots, k_c \leq n_x \\ 1 \leq s_1, \cdots, s_{a-c} \leq n_y}} \prod_{t=1}^{c} x_{k_t,j} \prod_{m=1}^{a-c} y_{s_m,j},$$

which allows the indexes k's and s's to be the same and thus is different from the U-statistics in (3.20). Xu et al. (2016) shows that the constructed V-statistics are biased *estimators of $\|\boldsymbol{\mu}-\boldsymbol{\nu}\|_a^a$, and $\mathcal{V}(a)$ and $\mathcal{V}(b)$ are asymptotically independent if $a+b$ is odd, but are asymptotically correlated if $a+b$ is even. The constructed U-statistics in this work extend the properties of those V-statistics such that $\mathcal{U}(a)$ in* (3.20) *is an unbiased estimator of $\|\boldsymbol{\mu}-\boldsymbol{\nu}\|_a^a$, and all $\mathcal{U}(a)$'s are asymptotically independent with each other. Given these nice statistical properties, it becomes easier to obtain the joint asymptotic distribution of the U-statistics, and then apply the adaptive test.*

3.4 Two-Sample Covariance Test

The U-statistics framework can be applied similarly to testing the equality of two covariance matrices. Suppose $\{\mathbf{x}_i\}_{i=1}^{n_x}$ and $\{\mathbf{y}_i\}_{i=1}^{n_y}$ are i.i.d. copies of two independent random vectors $\mathbf{x}=(x_1,\ldots,x_p)^\intercal$ and $\mathbf{y}=(y_1,\ldots,y_p)^\intercal$ respectively. Denote $\mathrm{E}(\mathbf{x})=\boldsymbol{\mu}=(\mu_1,\ldots,\mu_p)^\intercal$, $\mathrm{E}(\mathbf{y})=\boldsymbol{\nu}=(\nu_1,\ldots,\nu_p)^\intercal$; $\mathrm{cov}(\mathbf{x})=\boldsymbol{\Sigma}_x=\{\sigma_{x,j_1,j_2}:1\le j_1,j_2\le p\}$ and $\mathrm{cov}(\mathbf{y})=\boldsymbol{\Sigma}_y=\{\sigma_{y,j_1,j_2}:1\le j_1,j_2\le p\}$. Consider $H_0:\boldsymbol{\Sigma}_x=\boldsymbol{\Sigma}_y=\boldsymbol{\Sigma}=(\sigma_{j_1,j_2})_{p\times p}$. Given $1\le j_1,j_2\le p$, $1\le k_1\ne k_2\le n_x$, and $1\le s_1\ne s_2\le n_y$, $K_{j_1,j_2}(\mathbf{x}_{k_1},\mathbf{x}_{k_2},\mathbf{y}_{s_1},\mathbf{y}_{s_2})=(x_{k_1,j_1}x_{k_1,j_2}-x_{k_1,j_1}x_{k_2,j_2})-(y_{s_1,j_1}y_{s_1,j_2}-y_{s_1,j_1}y_{s_2,j_2})$ is a simple unbiased estimator of $\sigma_{x,j_1,j_2}-\sigma_{y,j_1,j_2}$. Therefore, for a finite positive integer a, we have the U-statistic

$$\mathcal{U}(a)=\sum_{1\le j_1,j_2\le p}\frac{1}{P_{2a}^{n_x}P_{2a}^{n_y}}\sum_{\substack{1\le k_{1,1}\ne k_{1,2}\ne\ldots\\ \ne k_{a,1}\ne k_{a,2}\le n_x}}\sum_{\substack{1\le s_{1,1}\ne s_{1,2}\ne\ldots\\ \ne s_{a,1}\ne s_{a,2}\le n_y}}\prod_{t=1}^{a}K_{j_1,j_2}(\mathbf{x}_{k_{t,1}},\mathbf{x}_{k_{t,2}},\mathbf{y}_{s_{t,1}},\mathbf{y}_{s_{t,2}}). \tag{3.23}$$

As in Remark III.1, another formulation of $\mathcal{U}(a)$ equivalent to (3.23) is

$$\begin{aligned}\mathcal{U}(a) = &\sum_{c=0}^{a}\sum_{b_1=0}^{c}\sum_{b_2=0}^{a-c}(-1)^{c-b_1+b_2} \sum_{1\le j_1,j_2\le p}\ \sum_{\substack{1\le i_1\neq\ldots\neq\\ i_{2c-b_1}\le n_x}}\ \sum_{\substack{1\le w_1\neq\ldots\neq\\ w_{2(a-c)-b_2}\le n_y}} \\ &C_{n_x,n_y,a,c,b_1,b_2}\times\prod_{k=1}^{b_1}(x_{i_k,j_1}x_{i_k,j_2})\prod_{s=b_1+1}^{c}x_{i_s,j_1}\prod_{t=c+1}^{2c-b_1}x_{i_t,j_2} \\ &\times\prod_{m=1}^{b_2}(y_{w_m,j_1}y_{w_m,j_2})\prod_{l=b_2+1}^{a-c}y_{w_l,j_1}\prod_{q=a-c+1}^{2(a-c)-b_2}y_{w_q,j_2},\end{aligned} \tag{3.24}$$

where $C_{n_x,n_y,c,b_1,b_2} = (P^{n_x}_{2c-b_1}P^{n_y}_{2(a-c)-b_2})^{-1}a!/\{b_1!(c-b_1)!b_2!(a-c-b_2)!\}$, and (3.24) shall be used in the theoretical developments.

We next present the asymptotic results of the constructed U-statistics under the null hypothesis. Here we assume the regularity Condition 3.4.1 or 3.4.1* that are assumed under H_0, where $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y = \mathbf{\Sigma} = (\sigma_{j_1,j_2})_{p\times p}$.

Condition 3.4.1 (Dependence Assumption: Mixing-Type)**.**

(1) $n, p \to \infty$, *and* $n_x/n \to \gamma \in (0,1)$.

(2) $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^8 < \infty$; $\lim_{p\to\infty}\min_{1\le j\le p}\mathrm{E}(x_j-\mu_j)^2 > 0$; $\lim_{p\to\infty}\max_{1\le j\le p}\mathrm{E}(y_j-\nu_j)^8 < \infty$; *and* $\lim_{p\to\infty}\min_{1\le j\le p}\mathrm{E}(y_j-\nu_j)^2 > 0$.

(3) $\{(x_{i,j}, i=1,\ldots,n) : 1\le j\le p\}$ *and* $\{(y_{i,j}, i=1,\ldots,n) : 1\le j\le p\}$ *are* α*-mixing with* $\alpha_x(s)\le C\delta_x^s$ *and* $\alpha_y(s)\le C\delta_y^s$, *where* $\delta_x,\delta_y\in(0,1)$ *and* C *is some constant.*

(4) For any finite integer a, $\sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a = \Theta(p^2)$.

Condition 3.4.1 (2) is similar to Condition 3.2.1. Condition 3.4.1 (3) assumes α-mixing on the two samples, which is similar to Condition 3.2.2. Condition 3.4.1 (4) is a regularity condition on the covariance structure, and it is naturally satisfied for even a, given Condition 3.4.1 (3).

Alternatively, we introduce another set of conditions similar to Condition 3.2.2*. We define some notation. Suppose $(z_1,\ldots,z_p)^\intercal \sim \mathcal{N}(\mathbf{0},\mathbf{\Sigma})$. Given indexes $1 \le$

$j_1, \ldots, j_t \le p$, define $\Pi^0_{j_1,\ldots,j_t} = \mathrm{E}(\prod_{k=1}^t z_{j_k})$. Moreover, we define $\Pi^x_{j_1,\ldots,j_t} = \mathrm{E}\{\prod_{k=1}^t (x_{j_k} - \mu_{j_k})\}$ and $\Pi^y_{j_1,\ldots,j_t} = \mathrm{E}\{\prod_{k=1}^t (y_{j_k} - \nu_{j_k})\}$. In addition, for given integers a and b, let $\mathbb{G}_{a,b}$ be a collection of tuples $\mathcal{G} = (g_1, g_2, \ldots, g_{4(a+b)-1}, g_{4(a+b)}) \in \{1, \ldots, 8\}^{4(a+b)}$, which satisfies that $g_{2t-1} \neq g_{2t}$ for $t = 1, \ldots, 2(a+b)$, and the number of g's equal to m is a for $m \in \{1, 2, 3, 4\}$ and is b for $m \in \{5, 6, 7, 8\}$. For any $\mathcal{G} \in \mathbb{G}_{a,b}$, we define $\mathbb{V}_{a,b,\mathcal{G}} = \sum_{1 \le j_1,\ldots,j_8 \le p} \prod_{t=1}^{2(a+b)} \sigma_{j_{g_{2t-1}}, j_{g_{2t}}}$, and let $S_{\mathcal{G}}$ denote the number of distinct sets among the $2(a+b)$ number of sets, $\{g_{2t-1}, g_{2t}\}$, for $t = 1, \ldots, 2(a+b)$, induced by $\mathcal{G}$. Note that generally $S_{\mathcal{G}} \ge 4$, and when $S_{\mathcal{G}} = 4$, by the symmetricity of j indexes, $\mathbb{V}_{a,b,\mathcal{G}} = \mathbb{V}_{a,b,0}$ where $\mathbb{V}_{a,b,0} := \sum_{1 \le j_1,\ldots,j_8 \le p} \sum_{1 \le j_1,\ldots,j_8 \le p} \sigma^a_{j_1,j_2} \sigma^a_{j_3,j_4} \sigma^b_{j_5,j_6} \sigma^b_{j_7,j_8}$.

Condition 3.4.1* (Alternative Dependence Assumption)**.**

(1) $n, p \to \infty$, and $n_x/n \to \gamma \in (0, 1)$.

(2) $\lim_{p\to\infty} \max_{1\le j\le p} \mathrm{E}(x_j - \mu_j)^8 < \infty$; $\lim_{p\to\infty} \min_{1\le j\le p} \mathrm{E}(x_j - \mu_j)^2 > 0$; $\lim_{p\to\infty} \max_{1\le j\le p} \mathrm{E}(y_j - \nu_j)^8 < \infty$; and $\lim_{p\to\infty} \min_{1\le j\le p} \mathrm{E}(y_j - \nu_j)^2 > 0$.

(3) For $t = 3, 4, 6, 8$, there exist constants $\kappa_{x,t}, \kappa_{y,t} \ge 1$ such that $\Pi^x_{j_1,\ldots,j_t} = \kappa_{x,t}\Pi^0_{j_1,\ldots,j_t}$ and $\Pi^y_{j_1,\ldots,j_t} = \kappa_{y,t}\Pi^0_{j_1,\ldots,j_t}$.

(4) For $a, b \in \{a_1, \ldots, a_m\}$, and any $\mathcal{G} \in \mathbb{G}_{a,b}$ define above, if $S_{\mathcal{G}} > 4$, we assume $\mathbb{V}_{a,b,\mathcal{G}} = o(1)\mathbb{V}_{a,b,0}$.

We note that Condition 3.4.1* (3) and (4) are alternative dependence assumptions to Condition 3.4.1 (3) and (4). Condition 3.4.1* (3) is an extension from Condition 3.2.2*, and is also satisfied when the distributions of $\mathbf{x}$ and $\mathbf{y}$ follow elliptical distributions Kan (2008). Condition 3.4.1* (4) implies some weak dependence structure in covariance matrix $\mathbf{\Sigma}$ that extends the moment assumption for second-order U-statistics in Li and Chen (2012) to U-statistics of general orders.

Theorem 3.4.1. *Assume Condition 3.4.1 or Condition 3.4.1*. Then under H_0, for finite integers $\{a_1, \ldots, a_m\}$, $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, where for*

$a \in \{a_1, \ldots, a_m\}$,

$$\begin{aligned}\sigma^2(a) &= \text{var}\{\mathcal{U}(a)\} \\ &\simeq \sum_{1\leq j_1,j_2,j_3,j_4\leq p} a!\Big\{\frac{1}{n_x}(\Pi^x_{j_1,j_2,j_3,j_4} - \sigma_{j_1,j_2}\sigma_{j_3,j_4}) + \frac{1}{n_y}(\Pi^y_{j_1,j_2,j_3,j_4} - \sigma_{j_1,j_2}\sigma_{j_3,j_4})\Big\}^a\end{aligned}$$

with $\Pi^x_{j_1,j_2,j_3,j_4} = \text{E}\{\prod_{t=1}^4(x_{1,j_t} - \mu_{j_t})\}$ *and* $\Pi^y_{j_1,j_2,j_3,j_4} = \text{E}\{\prod_{t=1}^4(y_{1,j_t} - \nu_{j_t})\}$.

Theorem 3.4.1 provides the asymptotic independence and joint normality of the finite-order U-statistics, which are similar to Theorems 3.2.1, 3.3.1 and 3.3.3. To further study the power of these finite-order U-statistics, we next consider the alternative hypotheses where $\mathbf{\Sigma}_x \neq \mathbf{\Sigma}_y$. Let $\mathbb{J}_0$ be the largest subset of $\{1, \ldots, p\}$ such that $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2} = \sigma_{j_1,j_2}$ for any $j_1, j_2 \in \mathbb{J}_0$. We then obtain the following theorem under the regularity conditions given in Section B.3.2.

Theorem 3.4.2. *Under Conditions B.3.1 and B.3.2, for finite integers* $\{a_1, \ldots, a_m\}$, $[\mathcal{U}(a_1) - \text{E}\{\mathcal{U}(a_1)\}]/\sigma(a_1), \ldots, [\mathcal{U}(a_m) - \text{E}\{\mathcal{U}(a_m)\}]/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, *where*

$$\sigma^2(a) = \text{var}\{\mathcal{U}(a)\} \simeq a!C_{\kappa,a} \sum_{j_1,j_2,j_3,j_4\in\mathbb{J}_0} \sigma^a_{j_1,j_2}\sigma^a_{j_3,j_4},$$

and $C_{\kappa,a} = \{(\kappa_x - 1)/n_x + (\kappa_y - 1)/n_y\}^a + 2(\kappa_x/n_x + \kappa_y/n_y)^a$ *with* κ_x *and* κ_y *given in Condition B.3.1.*

Given the asymptotic results under the alternatives, we next analyze the power of the finite-order U-statistics. By Theorem 3.4.2, the asymptotic power of $\mathcal{U}(a)$ depends on $\text{E}\{\mathcal{U}(a)\}/\sqrt{\text{var}\{\mathcal{U}(a)\}}$. Let $J_D = \{(j_1, j_2) : \sigma_{x,j_1,j_2} \neq \sigma_{y,j_1,j_2}, 1 \leq j_1, j_2 \leq p\}$, then $\text{E}\{\mathcal{U}(a)\} = \sum_{(j_1,j_2)\in J_D}(\sigma_{x,j_1,j_2} - \sigma_{y,j_1,j_2})^a$. Similarly to Section 3.2.2, to study the relationship between the sparsity level of $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ and the power of U-statistics, we consider the case where the non-zero differences between $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ are the same. Specifically, let $\sigma_{x,j_1,j_2} - \sigma_{y,j_1,j_2} = \rho$ for $(j_1, j_2) \in J_D$, and then $\text{E}\{\mathcal{U}(a)\} = |J_D|\rho^a$. Following the analysis in Section 3.2.2, we compare the ρ values needed by different

$\mathcal{U}(a)$'s to achieve $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ for a given constant M. In particular, for given integer a, suppose $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ is achieved when $\rho = \rho_a$. For any $a \neq b$, we compare $\mathcal{U}(a)$ and $\mathcal{U}(b)$ following Criterion III.1.

We use the following example as an illustration, where $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ satisfy the conditions of Theorem 3.4.2. Specifically, we assume that $\mathbf{\Sigma}_x = (\sigma_{x,j_1,j_2})_{p\times p}$ has the diagonal elements $\sigma_{x,j,j} = \nu^2$; and the off-diagonal elements $\sigma_{x,j_1,j_2} = h_{|j_1-j_2|} \in (0, \nu^2)$ with $h_{|j_1-j_2|} = \Theta(\nu^2)$ when $|j_1 - j_2| \leq s$, while $\sigma_{x,j_1,j_2} = 0$ when $|j_1 - j_2| > s$. This covers the moving average covariance structure of order s, and $\mathbf{\Sigma}_x$ is a banded matrix with bandwidth s. In addition, we assume the bandwidth $s = o(p)$ and $p - |\mathbb{J}_0| = o(p)$. By the definition of $\mathbb{J}_0$, the assumption $p - |\mathbb{J}_0| = o(p)$ implies that a large square sub-matrix of $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ are the same. For simplicity, we let $n_x = n_y$ with $n = n_x + n_y$, and a similar analysis can be applied when $n_x \neq n_y$. By Theorem 3.4.2, $\mathrm{var}\{\mathcal{U}(a)\} \simeq (n/2)^{-a} a! \{2\kappa_1^a + \kappa_2^a\}\{p\nu^{2a} + 2\sum_{t=1}^{s} h_t^a (p-t)\}^2$, where $\kappa_1 = \kappa_x + \kappa_y$ and $\kappa_2 = \kappa_x + \kappa_y - 2$. Therefore we know for given finite integer a, $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ holds when $\rho = \rho_a$ defined as

$$\rho_a = \frac{(a!)^{\frac{1}{2a}} \sqrt{\kappa_1}\nu}{(n/2)^{1/2}} \Big(\frac{Mp}{|J_D|}\Big)^{1/a} \Big\{2 + \Big(\frac{\kappa_2}{\kappa_1}\Big)^a\Big\}^{\frac{1}{2a}} \Big\{1 + 2\sum_{t=1}^{s} \Big(\frac{h_t}{\nu^2}\Big)^a \Big(1 - \frac{t}{p}\Big)\Big\}^{\frac{1}{a}}.$$

We next compare the ρ_a's and obtain the following proposition.

Proposition 3.4.1. *There exists $\mathbb{D}_0$ that only depends on the given $\kappa_x, \kappa_y, \nu^2, s$, and $h_t, t = 1, \ldots, s$, and satisfies $\mathbb{D}_0 = \Theta(1/s^2)$ such that*

(i) When $|J_D| \geq Mp/\sqrt{\mathbb{D}_0}$, the minimum of ρ_a is achieved at $a_0 = 1$.

(ii) When $|J_D| < Mp/\sqrt{\mathbb{D}_0}$, the minimum of ρ_a is achieved at some a_0, which increases as $Mp/|J_D|$ increases.

Proposition 3.4.1 is similar to Propositions 3.2.1 and 3.3.1. Following the analysis in Section 3.2.2, Proposition 3.4.1 shows that when the difference $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ is "very" dense with $|J_D| \geq Mp/\sqrt{\mathbb{D}_0}$, $\mathcal{U}(1)$ is the most powerful U-statistic; when $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ becomes sparser as $Mp/|J_D|$ decreases, a higher order U-statistic is more powerful; when the $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ is "moderately" dense or sparse, a U-statistic of finite order $a_0 > 1$ would be the most powerful one.

The power analysis above shows that the power of the U-statistics varies when the alternative changes. To maintain high power across different alternatives, we can develop an adaptive testing procedure similar to that in Section 3.2.3. Given the asymptotic independence in Theorem 3.4.1, an adaptive testing procedure using the constructed $\mathcal{U}(a)$'s is valid with the type I error asymptotically controlled. Also, the adaptive test achieves high power by combining the U-statistics as discussed in Section 3.2.3.

We provide simulation studies on two-sample covariance testing in the Appendix Section B.7.3. By the simulations, we first find that the type I errors of the U statistics and the adaptive test are well controlled under H_0. This verifies the theoretical results in Theorem 3.4.2. Second, similarly to the one-sample covariance testing, we find that generally when the difference $\mathbf{\Sigma}_x - \mathbf{\Sigma}_y$ is sparser, a U-statistic of higher order is more powerful, and vice versa. Moreover, under moderately sparse/dense alternatives, $\mathcal{U}(a_0)$ with $a_0 > 1$ could achieve the highest power. The results are consistent with Proposition 3.4.1. Third, we compare the proposed adaptive test with existing methods in literature including Schott (2007), Srivastava and Yanagihara (2010), Li and Chen (2012), and Cai et al. (2013). We find that the proposed adaptive testing procedure maintains high power across various alternatives.

Remark III.6. *Similarly to Section 3.2, we can let $\mathcal{U}(\infty)$ be the maximum-type test statistic in Cai et al. (2013), and expect that the result similar to Theorem 3.2.3 holds under certain regularity conditions. However, as the dependence structure of two-sample covariance matrices is more complicated than the one-sample case, it is more challenging to establish the asymptotic joint distribution of $\mathcal{U}(\infty)$ and finite-*

order U-statistics. We leave this interesting problem for future study, while find in simulations that the performance of $\mathcal{U}(\infty)$ is similar to high-order U-statistics $\mathcal{U}(a)$'s.

3.5 Testing Coefficients in Generalized Linear Models

In this section, we consider the Example III.3 of generalized linear models (on Page 59) to show that the proposed framework can be extended to other testing problems. Similarly to the results in Section 3.3, we show that the constructed U-statistics are asymptotically independent and normally distributed, and also establish the power analysis results of the U-statistics. Recently, Wu et al. (2019) also discussed the adaptive testing of generalized linear model. But similarly to Xu et al. (2016), Wu et al. (2019) is under the framework of a family of von Mises V-statistics, and thus is different from the current chapter as discussed in Remark III.5. Moreover, the current work provides the theoretical power analysis while Wu et al. (2019) did not.

Condition 3.5.1.

(1) There exists constant B such that $B^{-1} \le \lambda_{\min}(\mathbf{\Sigma}) \le \lambda_{\max}(\mathbf{\Sigma}) \le B$, where $\lambda_{\min}(\mathbf{\Sigma})$ and $\lambda_{\max}(\mathbf{\Sigma})$ denote the minimum and maximum eigenvalues of the covariance matrix $\mathbf{\Sigma}$; and all correlations are bounded away from -1 and 1, i.e., $\max_{1\le j_1\neq j_2\le p} |\sigma_{j_1,j_2}|/(\sigma_{j_1,j_2}\sigma_{j_2,j_2})^{1/2} < 1-\eta$ for some $\eta > 0$.

(2) $\log p = o(n^{1/4})$ and $\max_{1\le j\le p} \mathrm{E}[\exp\{h(S_j - \mathrm{E}(S_j))^2\}] < \infty$, for $h \in [-M, M]$, where M is a positive constant.

(3) Similarly to Condition 3.2.2, $\{(S_{i,j}, i = 1\dots, n) : 1 \le j \le p\}$ is α-mixing with $\alpha_S(s) \le C\delta^s$, where $\delta \in (0,1)$ and C is some constant. In addition, for finite integer a, $\sum_{j_1,j_2=1}^{p} \sigma_{j_1,j_2}^a = \Theta(p)$.

Theorem 3.5.1. *Under Condition 3.5.1 and H_0: $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, for any finite integers $(a_1,\dots,a_m)$, as $n,p \to \infty$, $[\mathcal{U}(a_1)/\sigma(a_1),\dots,\mathcal{U}(a_m)/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, where*

$\sigma^2(a) = \sum_{i=1}^p \sum_{j=1}^p \sigma_{i,j}^a / P_a^n$, *which is of order* $\Theta(pn^{-a})$. *Besides,* $P(n\mathcal{U}(\infty) - \tau_p \leq u) \to \exp\{-\pi^{-1/2}\exp(-u/2)\}$, $\forall u \in \mathbb{R}$, *where* $\tau_p = 2\log p - \log\log p$. *In addition, for any finite integer* a, $\{\mathcal{U}(a)/\sigma(a)\}$ *and* $\{n\mathcal{U}(\infty) - \tau_p\}$ *are asymptotically independent.*

Next we compare the power of $\mathcal{U}(a)$'s under alternatives with different sparsity levels. Similarly to the mean testing problems, we consider the alternative $\mathcal{E}_A = \{\mathrm{E}(S_j) = \rho > 0 \text{ for } j = 1, \ldots, k_0; \mathrm{E}(S_j) = 0 \text{ for } j = k_0 + 1, \cdots, p\}$, where k_0 denotes the number of nonzero entries.

Theorem 3.5.2. *Assume Condition 3.5.1 and* $k_0 = o(p)$. *For any finite integers* $\{a_1, \ldots, a_m\}$, *if* ρ *in* $\mathcal{E}_A$ *satisfies* $\rho = O(k_0^{-1/a_t} p^{1/(2a_t)} n^{-1/2})$ *for* $t = 1, \ldots, m$, *then* $[\mathcal{U}(a_1) - \mathrm{E}\{\mathcal{U}(a_1)\}]/\sigma(a_1), \ldots, [\mathcal{U}(a_m) - \mathrm{E}\{\mathcal{U}(a_m)\}]/\sigma(a_m)]^\intercal \xrightarrow{D} \mathcal{N}(0, I_m)$, *as* $n, p \to \infty$. *In addition,* $\mathrm{E}[\mathcal{U}(a)] = \|\mathcal{E}_A\|_a^a = k_0\rho^a$ *and*

$$\sigma^2(a) \simeq \sum_{j_1=k_0+1}^{p} \sum_{j_2=k_0+1}^{p} a!\sigma_{j_1,j_2}^a / P_a^n,$$

which is $\Theta(a!pn^{-a})$.

Theorem 3.5.2 shows that under the considered local alternatives, the asymptotic power of $\mathcal{U}(a)$ mainly depends on $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}}$.

Therefore, for a given constant $M > 0$, if $\rho = \rho_a$ defined as $\rho_a = M^{1/a}k_0^{-1/a}a!^{1/(2a)} \times (\sum_{j_1=k_0+1}^p \sum_{j_2=k_0+1}^p \sigma_{j_1,j_2}^a)^{1/(2a)} \times n^{-1/2}$, we know that different $\mathcal{U}(a)$'s asymptotically have the same power. For illustration, we further assume that $\sigma_{j,j} = 1$ when $j \in \{k_0 + 1, \ldots, p\}$, and $\sigma_{j_1,j_2} = 0$ when $j_1 \neq j_2 \in \{k_0 + 1, \ldots, p\}$, then

$$\rho_a \simeq (M\sqrt{p}/k_0)^{\frac{1}{a}} a!^{\frac{1}{2a}} n^{-\frac{1}{2}}. \tag{3.25}$$

Therefore, following the analysis in Section 3.3, to find the "best" $\mathcal{U}(a)$, it suffices to find the order, denoted by a_0, that gives the smallest ρ_a value in (3.25). Since (3.25) is only different from (3.22) by a constant that does not depend on the order a, Proposition 3.3.1 still holds. Consider $a_0 \geq 1$ as specified in Proposition 3.3.1; then, similar to results in the two-sample mean testing, we know when $k_0 \geq \sqrt{Mp}$, $a_0 = 1$ and $\mathcal{U}(1)$ is "better" than $\mathcal{U}(\infty)$; when $k_0 < C_1\sqrt{p}/\log^{a_0/2} p$ for some C_1, $\mathcal{U}(\infty)$ is the "best"; and when $C_2\sqrt{p}/\log^{a_0/2} p < k_0 < \sqrt{Mp}$ for some C_2, $\mathcal{U}(a_0)$ is the "best". In addition, given the similar results obtained in Theorem 3.5.1 and power analysis, we can also develop adaptive testing procedure similar to that in Section 3.2.3.

Remark III.7. *More generally, if the generalized linear model also has covariates* $\mathbf{z}$ *that we want to adjust for, the corresponding generalized linear model becomes* $\mathrm{E}(y|\mathbf{x}) = g^{-1}(\mathbf{x}^\intercal\boldsymbol{\beta} + \mathbf{z}^\intercal\boldsymbol{\alpha})$, *where* $\boldsymbol{\alpha}$ *denote the regression coefficients for* $\mathbf{z}$. *To test* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ *v.s.* $H_A : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, *we can replace* $\mu_{0,j}$ *by* $\hat{\mu}_{0,j} = g^{-1}(\mathbf{x}_i^\intercal\boldsymbol{\beta}_0 + \mathbf{z}_i^\intercal\hat{\boldsymbol{\alpha}})$ *where* $\hat{\boldsymbol{\alpha}}$ *is an estimator of* $\boldsymbol{\alpha}$. *For instance, when* $\mathbf{z}$ *is low dimensional, we can take* $\hat{\boldsymbol{\alpha}}$ *as the maximum likelihood estimator under* H_0. *Then similar conclusion to Theorem 3.5.1 can be derived under certain regularity conditions. We present simulation studies on generalized linear model in the Appendix Section B.7.2 to illustrate the good performance of the U-statistics and we leave the details of theoretical developments with nuisance parameters for future study.*

3.6 Discussion

There are several possible extensions of the U-statistics framework in this chapter. First, by our current proof, the convergence rate in Theorem 3.2.3 is bounded by $O(\log^{-1/2} p)$, which is an upper bound and not sharp. From our extensive simulations, we find that the type I error rate of the adaptive testing is well-controlled with a relatively small p, e.g., $p = 50$. We might obtain a shaper bound of the convergence rate, but more refined concentration property of the high-dimensional and high-order U-statistics is needed. Second, the proposed framework requires that the elements in the parameter set $\mathcal{E}$ have unbiased estimates. When we can not obtain unbiased estimates easily, e.g., for the precision matrix, the proposed construction may not follow directly. Nevertheless we may use "nearly" unbiased estimators to construct "U-statistics" for hypothesis testing, such as the "nearly" unbiased estimator of the precision matrix proposed in Xia et al. (2015); the main challenge is then to control the accumulative bias over the parameters under high-dimensions. Third, this chapter discusses the examples where the elements in $\mathcal{E}$ are comparable. When the parameters in $\mathcal{E}$ are not comparable, such as $\mathcal{E}$ containing both means and covariances parameters, the construction of U-statistics still follows but the theoretical derivation may require a careful case-by-case examination. Fourth, the construction of the U-statistics treats the parameters in $\mathcal{E}$ with equal weight. More generally, we could assign different weights to different parameter estimators. For instance, standardizing the data is one example of assigning different weights. As inappropriate weight assignments could lead to power loss, when the truth is unknown, how to effectively assign weights to maximize the test power is an interesting research question. We shall discuss these extensions in the future as a significant amount of additional work is still needed.

In addition to the examples in this chapter, the proposed U-statistics framework can be applied to other high-dimensional hypothesis testing problems. For example, it can be applied to testing the block-diagonality of a covariance matrix, whose theoretical analysis would be similar to the considered one sample and two sample covariance testing problems. It can also be used to test high-dimensional regression coefficients in complex regression models other than the generalized linear models, following a similar construction based on the score functions. A key step is then to characterize the impact of nuisance parameters that are estimated under the null hypothesis, and challenges arise especially when the nuisance parameters are high-dimensional. Such interesting extensions will be further explored in our follow-up studies.

CHAPTER IV

Importance Sampling of Rare-Event Probabilities

This section develops an efficient Monte Carlo method to estimate the tail probabilities of the ratio of the largest eigenvalue to the trace of the Wishart matrix, which plays an important role in multivariate data analysis. The estimator is constructed based on a change-of-measure technique and it is proved to be asymptotically efficient for both the real and complex Wishart matrices. This chapter is organized as follows. In Section 4.1, we introduce the background and the set-up of the problem. In Section 4.2, we propose our importance sampling estimator and establish its asymptotic efficiency in Theorem 4.2.1. In Section 4.3, we present simulation studies to show the improved performance of the proposed method over existing approaches based on asymptotic approximations, especially when estimating probabilities of rare events. We discuss the possibility of generalizing the result to the ratio of the sum of the largest k eigenvalues to the trace of a Wishart matrix in Section 4.4. The proof of Theorem 4.2.1 is given in Section 4.5.

4.1 Introduction

Consider n independent and identically distributed (i.i.d.) p-dimensional observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from a real or complex valued Gaussian distribution with mean zero and covariance matrix $\Sigma = \sigma^2 I_p$. Here σ^2 is an unknown scaling factor and I_p

is the $p \times p$ identity matrix. Define the $n \times p$ data matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$, and assume $\lambda_1 \geq \cdots \geq \lambda_p$ are the ordered real eigenvalues of the sample covariance matrix $\hat{\Sigma} = \mathbf{X}^H\mathbf{X}/n$, where ^{H} denotes the conjugate transpose. Note that if $p > n$, the last $p - n$ of the λs are zero. Let $U_{n,p}$ be the ratio of the largest eigenvalue to the trace, viz.,

$$U_{n,p} = \frac{\lambda_1}{(\lambda_1 + \cdots + \lambda_p)/\min(n,p)}. \tag{4.1}$$

We are interested in estimating the rare-event tail probability $\alpha_{n,p}(x) = \Pr\left(U_{n,p} > x\right)$, where x is some constant such that $\alpha_{n,p}(x)$ is small. Estimating rare-event tail probabilities is often of interest in multivariate data analysis. For instance, in multiple testing problems, it is often needed to evaluate very small p-values for individual test statistics to control the overall false-positive error rate.

The random variable $U_{n,p}$ plays an important role in multivariate statistics when testing the covariance structure. For instance, it has been used to test for equality of the population covariance to a scaled identity matrix, viz.,

$$\mathcal{H}_0 : \Sigma = \sigma^2 I_p \quad \text{vs.} \quad \mathcal{H}_1 : \Sigma \neq \sigma^2 I_p$$

with σ^2 unknown, i.e., the so-called sphericity test; see, e.g., Muirhead (2009). The test statistic $U_{n,p}$ does not depend on the unknown variance parameter σ^2 and has high detection power against alternative covariance matrices with a low-rank perturbation of the null $\sigma^2 I_p$. In particular, under the alternative of rank-1 perturbation with $\Sigma = hh^\top + \sigma^2 I_p$ for some unknown $h \in \mathbb{R}^p$ and σ^2, the likelihood ratio test statistic $\mathcal{L}_n = \sup_{h,\sigma^2} f_1(X; h, \sigma^2)/\sup_{\sigma^2} f_0(X; \sigma^2)$ can be written as a monotone function of $U_{n,p}$ and therefore $\alpha_{n,p}(x)$ corresponds to the p-value (see, e.g., Muirhead, 2009; Bianchi et al., 2011). Please refer to Krzanowski (2000), Muirhead (2009), and Paul and Aue (2014) for more discussion and many other applications.

The exact distribution of $U_{n,p}$ is difficult to compute, especially when estimating rare-event tail probabilities. Note that $\mathbf{X}^H\mathbf{X}/(n\sigma^2)$ follows a Wishart distribution $\mathcal{W}_{\beta,p}(n, I_p/n)$, with $\beta = 1$ for real Gaussian and $\beta = 2$ for complex Gaussian. So the distribution of $U_{n,p}$ corresponds to that of the ratio of the largest eigenvalue to the trace of a $\mathcal{W}_{\beta,p}(n, I_p/n)$. However, this distribution is nonstandard and exact formulas based on it typically involve high-dimensional integrals or inverses of Laplace transforms. Numerical evaluation has been studied in Davis (1972), Schuurmann et al. (1973), Kuriki and Takemura (2001), Kortun et al. (2012), Wei et al. (2012), and Chiani (2014). But for high-dimensional data with large p, the computation becomes more challenging, which is notably the case when $\alpha_{n,p}(x)$ is small, due to the additional computational cost to control the relative estimation error of $\alpha_{n,p}(x)$.

The asymptotic distribution of $U_{n,p}$ with p and n both going to infinity has also been studied in the literature. It is known that $U_{n,p}$ asymptotically behaves similarly to the largest eigenvalue λ_1, whose limiting distribution has been studied in Johansson (2000) and Johnstone (2001), and $U_{n,p}$ also asymptotically follows the Tracy–Widom distribution (see, e.g., Bianchi et al., 2011; Nadler, 2011). That is,

$$\Pr\left(\frac{U_{n,p} - \mu_{n,p}}{\sigma_{n,p}} > x\right) \to 1 - \mathcal{TW}_\beta(x), \tag{4.2}$$

where $\mathcal{TW}_\beta$ denotes the Tracy–Widom distribution of order β, with $\beta = 1$ or 2 for real and complex valued observations, respectively. In particular, for real-valued observations, the centering and scaling constants

$$\begin{aligned} \mu_{n,p} &= \frac{1}{n}\left(\sqrt{n-\frac{1}{2}} + \sqrt{p-\frac{1}{2}}\right)^2 \\ \sigma_{n,p} &= \frac{1}{n}\left(\sqrt{n-\frac{1}{2}} + \sqrt{p-\frac{1}{2}}\right)\left(\frac{1}{\sqrt{n-1/2}} + \frac{1}{\sqrt{p-1/2}}\right)^{1/3} \end{aligned} \tag{4.3}$$

lead to a convergence rate of the order $O\{\min(n,p)^{-2/3}\}$; see Ma (2012). For the complex case, similar expressions can be found in Karoui (2006). Nadler (2011) studied the accuracy of the Tracy–Widom approximation for finite values of n and p. He found that the approximation may be inaccurate for small and even moderate values of p when n is large. Therefore, he proposed a correction term to improve the approximation result, which is derived using the Fredholm determinant representation, and he showed that the approximation rate is $o\{\min(n,p)^{-2/3}\}$ when X follows a *complex* Gaussian distribution. In the *real* Gaussian case, which is of interest in many statistical applications, Nadler (2011) conjectured that the result also holds. The calculation of the correction term in Nadler (2011) depends on the second derivative of the non-standard Tracy–Widom distribution, which usually involves a numerical discretization scheme.

Another limitation of the existing methods is that they may become less efficient when estimating small tail probabilities of rare events. This chapter aims to address this rare-event estimation problem. In particular, we propose an efficient Monte Carlo method to estimate the exact tail probability of $U_{n,p}$ by utilizing importance sampling. The latter is a commonly used tool to reduce Monte Carlo variance and it has been found helpful to estimate small tail probabilities, especially when the event is rare, in a wide variety of stochastic systems with both light-tailed and heavy-tailed distributions (see, e.g., Siegmund, 1976; Asmussen and Kroese, 2006; Dupuis et al., 2007; Asmussen and Glynn, 2007; Blanchet and Glynn, 2008; Liu and Xu, 2014a,b; Xu et al., 2014).

An importance sampling algorithm needs to construct an alternative sampling measure (a change of measure) under which the eigenvalues are sampled. Note that it is necessary to normalize the estimator with a Radon–Nikodym derivative to ensure an unbiased estimate. Ideally, one develops a sampling measure so that the event of interest is no longer rare under the sampling measure. The challenge is of course the construction of an appropriate sampling measure, and one common heuristic is to utilize a sampling measure that approximates the conditional distribution of $U_{n,p}$ given the event $\{U_{n,p} > x\}$. This chapter proposes a change of measure Q that asymptotically approximates the conditional measure $\Pr(\cdot \mid U_{n,p} > x)$. We carry out a rigorous analysis of the proposed estimator for $U_{n,p}$ and show that it is asymptotically efficient. Simulation studies show that the proposed method outperforms existing approximation approaches, especially when estimating probabilities of rare events.

4.2 Importance Sampling Estimation

For ease of discussion, we consider the setting $p \leq n$, $p \to \infty$ and $n \to \infty$. When $p > n$, the algorithm and theory are essentially the same up to switching labels of p and n, which is explained in Remark IV.4. We use the notation β to denote the real Wishart Matrix ($\beta = 1$) and complex Wishart matrix ($\beta = 2$). Since $U_{n,p} = p\lambda_1/(\lambda_1 + \cdots + \lambda_p)$ is invariant to σ^2, the analysis does not depend on the specific values of σ^2, and we take σ^2 as follows in order to simplify the notation and unify the real and complex cases under the same representation, as specified in Eq. (4.4) below:

a) When $\beta = 1$, we assume that $\sigma^2 = 1$. That is, the entries of $\mathbf{X}$ are i.i.d. $\mathcal{N}(0, 1)$, and $\lambda_1, \ldots, \lambda_p$ are the ordered eigenvalues of $\mathbf{X}^\top \mathbf{X}/n$.

b) When $\beta = 2$, we assume $\sigma^2 = 2$. We consider the circularly symmetric Gaussian random variable (Tse and Viswanath, 2005), and we write $X = Y + iZ \sim \mathcal{CN}(0, \sigma^2)$ when Y and Z are i.i.d. $\mathcal{N}(0, \sigma^2/2)$. In the following, we assume that the entries of $\mathbf{X}$ are i.i.d. $\mathcal{CN}(0, 2)$, and that $\lambda_1, \ldots, \lambda_p$ are the ordered eigenvalues of $\mathbf{X}^H \mathbf{X}/n$.

As mentioned, e.g., in (Dumitriu and Edelman, 2002), the p eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ are distributed with probability density function

$$f_{n,p,\beta}(\lambda) = C_{n,p,\beta} \prod_{i<j}^{p} |\lambda_i - \lambda_j|^{\beta} \prod_{i=1}^{p} \lambda_i^{\beta(n-p+1)/2-1} e^{-n(\lambda_1+\cdots+\lambda_p)/2}, \tag{4.4}$$

when $\beta \in \{1, 2\}$, where $C_{n,p,\beta}$ is a normalizing constant given by

$$C_{n,p,\beta} = p! \left(\frac{n}{2}\right)^{\beta np/2} \prod_{j=1}^{p} \frac{\Gamma(1+\beta/2)}{\Gamma(1+\beta j/2)\Gamma\{\beta(n-p+j)/2\}}.$$

Then the target probability $\alpha_{n,p}(x) = \Pr(U_{n,p} > x)$ can be written as

$$\alpha_{n,p}(x) = \int_{\lambda_1 \geq \cdots \geq \lambda_p \geq 0} \mathbf{1}(U_{n,p} > x) f_{n,p,\beta}(\lambda_1, \ldots, \lambda_p) d\lambda_1 \cdots d\lambda_p,$$

where $\mathbf{1}$ is the indicator function. As discussed in the Introduction, direct evaluation of the above p-dimensional integral is computationally challenging, especially when p is relatively large.

This work aims to design an efficient Monte Carlo method to estimate $\alpha_{n,p}(x)$. We first introduce some computational concepts from the rare-event analysis literature, which helps to evaluate the computation efficiency of a Monte Carlo estimator.

Consider an estimator $L_{n,p}(x)$ of a rare-event probability $\alpha_{n,p}(x)$, which goes to 0 as $n \to \infty$. We simulate N i.i.d. copies of $L_{n,p}(x)$, say $L_{n,p}^{(1)}(x), \ldots, L_{n,p}^{(N)}(x)$, and obtain the average estimator $\bar{L}_{n,p}(x) = \{L_{n,p}^{(1)}(x) + \cdots + L_{n,p}^{(N)}(x)\}/N$. We want to control the relative error $|\bar{L}_{n,p}(x) - \alpha_{n,p}(x)|/\alpha_{n,p}(x)$ such that for some prescribed $\varepsilon, \delta \in (0, \infty)$,

$$\Pr\{|\bar{L}_{n,p}(x) - \alpha_{n,p}(x)|/\alpha_{n,p}(x) > \varepsilon\} < \delta.$$

Consider the direct Monte Carlo estimator as an example. The direct Monte Carlo directly generates samples from the density (4.4) and uses $L_{n,p}(x) = \mathbf{1}(U_{n,p} > x)$.

So in each simulation we have a Bernoulli variable with mean $\alpha_{n,p}(x)$. According to the Central Limit theorem, the direct Monte Carlo simulation requires $N = \Theta\{\varepsilon^{-2}\delta^{-1}\alpha_{n,p}(x)^{-1}\}$ i.i.d. replicates to achieve the above accuracy. This implies that the direct Monte Carlo method becomes inefficient and even infeasible as $\alpha_{n,p}(x) \to 0$.

A more efficient estimator is the asymptotically efficient estimator (see, e.g., Siegmund, 1976; Asmussen and Kroese, 2006). An unbiased estimator $L_{n,p}(x)$ of $\alpha_{n,p}(x)$ is called asymptotically efficient if

$$\liminf_{n\to\infty} \ln[\mathrm{var}\{L_{n,p}(x)\}]/\ln\{\alpha_{n,p}(x)^2\} \geq 1. \tag{4.5}$$

Note that (4.5) is equivalent to

$$\limsup_{n\to\infty} \mathrm{var}\{L_{n,p}(x)\}/\alpha_{n,p}(x)^{2-\eta} = 0, \tag{4.6}$$

for any $\eta > 0$. In addition, since $\mathrm{E}(L_{n,p}^2) \geq \mathrm{var}\{L_{n,p}(x)\}$ and

$$\limsup_{n\to\infty} \ln\{\mathrm{E}(L_{n,p}^2)\}/\ln\{\alpha_{n,p}(x)^2\} \leq 1$$

by Hölder's inequality, (4.5) is also equivalent to

$$\lim_{n\to\infty} \ln\{\mathrm{E}(L_{n,p}^2)\}/\ln\{\alpha_{n,p}(x)^2\} = 1.$$

When $L_{n,p}(x)$ is asymptotically efficient, by Chebyshev's inequality,

$$\Pr\{|\bar{L}_{n,p}(x) - \alpha_{n,p}(x)|/\alpha_{n,p}(x) > \varepsilon\} \leq \mathrm{var}\{L_{n,p}(x)\}/\{N\alpha_{n,p}(x)^2\epsilon^2\},$$

and therefore (4.6) implies that we only need $N = O\{\varepsilon^{-2}\delta^{-1}\alpha_{n,p}(x)^{-\eta}\}$, for any $\eta > 0$, i.i.d. replicates of $L_{n,p}(x)$. Compared with the direct Monte Carlo simulation, efficient estimation substantially reduces the computational cost, especially when $\alpha_{n,p}(x)$ is small.

To construct an asymptotically efficient estimator, we use the importance sampling technique, which is an often used method for variance reduction of a Monte Carlo estimator. We use P to denote the probability measure of the eigenvalues $\lambda_1, \ldots, \lambda_p$. The importance sampling estimator is constructed based on the identity

$$\Pr(U_{n,p} > x) = \mathrm{E}\{\mathbf{1}(U_{n,p} > x)\} = \mathrm{E}_Q\left\{\mathbf{1}(U_{n,p} > x)\, dP/dQ\right\},$$

where Q is a probability measure such that the Radon–Nikodym derivative dP/dQ is well defined on the set $\{U_{n,p} > x\}$, and we use E and E_Q to denote the expectations under the measures P and Q, respectively. Let $f_{n,p}^Q$ be the density function of the eigenvalues $\lambda_1, \ldots, \lambda_p$ under the change of measure Q. Then, the random variable defined by

$$L_{n,p} = \mathbf{1}(U_{n,p} > x) f_{n,p}(\lambda_1, \ldots, \lambda_p)/f_{n,p}^Q(\lambda_1, \ldots, \lambda_p)$$

is an unbiased estimator of $\alpha_{n,p}(x)$ under the measure Q. Therefore, to have $L_{n,p}$ asymptotically efficient, we only need to choose a change of measure Q such that

$$\liminf_{n\to\infty} \frac{1}{|2\ln \alpha_{n,p}(x)|} \,|\ln \mathrm{E}_Q\{\mathbf{1}(U_{n,p} > x) f_{n,p}(\lambda_1, \ldots, \lambda_p)^2/f_{n,p}^Q(\lambda_1, \ldots, \lambda_p)^2\}\,| \geq 1. \tag{4.7}$$

To gain insight into the requirement (4.7), we consider some examples. First consider the direct Monte Carlo with $f_{n,p}^Q = f_{n,p}$; the right-hand side of (4.7) then equals 1/2 which is smaller than 1. On the other hand, consider Q to be the conditional probability measure given $U_{n,p} > x$, i.e., $f_{n,p}^Q(\cdot) = f_{n,p}(\cdot)\mathbf{1}(U_{n,p} > x)/\alpha_{n,p}(x)$; then the right-hand side of (4.7) is exactly 1. Note that this change of measure is of no practical use since $L_{n,p}$ depends on the unknown $\alpha_{n,p}(x)$. But if we can find a measure Q that is a good approximation of the conditional probability measure given $U_{n,p} > x$, we would expect (4.7) to hold and the corresponding estimator $L_{n,p}$ to be efficient.

In other words, the asymptotic efficiency criterion requires the change of measure Q to be a good approximation of the conditional distribution of interest.

Following the above argument, we construct the change of measure Q as follows, which is motivated by a recent study of Jiang et al. Jiang, Leder, and Xu (2017). These authors studied the tail probability of the largest eigenvalue, i.e., $\Pr(\lambda_1 > px)$ with $p > n$ and proposed a change of measure that approximates the conditional probability measure given $\lambda_1 > px$ in total variation when $p \gg n$. It is known that the asymptotic behaviors of λ_1 and $U_{n,p}$ are closely related. We therefore adapt the change of measure to the current problem of estimating $U_{n,p}$. However, we would like to clarify that the problem of estimating $U_{n,p}$ is different from that in Jiang et al. (2017) in terms of both theoretical justification and computational implementation, which is further discussed in Remark IV.3.

Specifically, we propose the following importance sampling estimator.

Algorithm IV.1. *Every iteration in the algorithm contains three steps, as follows:*

Step 1. We use the matrix representation of the β-Laguerre ensemble in Dumitriu and Edelman (2002), and generate the matrix $\mathbf{L}_{n-1,p-1,\beta} = \mathbf{B}_{n-1,p-1,\beta}\mathbf{B}_{n-1,p-1,\beta}^{\top}$, where $\mathbf{B}_{n-1,p-1,\beta}$ is a bidiagonal matrix defined by

$$\mathbf{B}_{n-1,p-1,\beta} = \begin{pmatrix} \chi_{\beta(n-1)} & & & \\ \chi_{\beta(p-2)} & \chi_{\beta(n-2)} & & \\ & \ddots & \ddots & \\ & & \chi_{\beta} & \chi_{\beta\{n-(p-1)\}} \end{pmatrix}_{(p-1)\times(p-1)}.$$

The notation χ_a denotes the square root of the chi-square distribution with a degrees of freedom, and the diagonal and sub-diagonal elements of $\mathbf{B}_{n-1,p-1,\beta}$

are generated independently. We then compute the corresponding ordered eigenvalues of $\mathbf{L}_{n-1,p-1,\beta}/n$*, denoted by* $\lambda_2 \geq \cdots \geq \lambda_p$.

Step 2. *Conditional on* $\lambda_2, \ldots, \lambda_p$*, we sample* λ_1 *from an exponential distribution with density*

$$f(\lambda_1) = nre^{-nr(\lambda_1 - \tilde{x} \vee \lambda_2)} \times \mathbf{1}\left(\lambda_1 > \tilde{x} \vee \lambda_2\right), \tag{4.8}$$

where $a \vee b = \max(a, b)$ *and* r *is a rate function such that*

$$r = \frac{1}{2} - \beta\gamma \int \frac{1}{\beta x - y}\, d\sigma_\beta(y) - \frac{1-\gamma}{2x} \tag{4.9}$$

with $\gamma = p/n$ *and* σ_β *denotes the probability distribution function of the Marchenko–Pastur law such that*

$$\sigma_\beta(ds) = (\beta \times 2\pi\gamma s)^{-1}\sqrt{(s - s_*)(s^* - s)}\,\mathbf{1}(s \in [s_*, s^*])ds \tag{4.10}$$

with $s^* = \beta(\sqrt{\gamma}+1)^2$ *and* $s_* = \beta(\sqrt{\gamma}-1)^2$*, and* $\tilde{x}$ *is a constant depending on* n*,* p*,* β *and* x *such that*

$$\tilde{x} = x \operatorname{tr}(\mathbf{L}_{n-1,p-1,\beta}/n)/(p - x).$$

Step 3. *Based on the collected values* $\lambda_1 \geq \cdots \geq \lambda_p$*, a corresponding importance sampling estimate can be computed as in* (4.12) *below and the value of the estimate is saved.*

The three steps above are repeated at every iteration. After the last iteration, the saved sampling estimates from all iterations are averaged to give an unbiased estimate of $\alpha_{n,p}(x)$.

Now we detail how the importance sampling estimate (4.12) is computed at every iteration of the algorithm. Let Q be the measure induced by combining the above two-step sampling procedure. From Dumitriu and Edelman (2002), under the change of measure Q, the density of $(\lambda_2^*, \ldots, \lambda_p^*) = n(\lambda_2, \ldots, \lambda_p)/(n-1)$ is

$$f_{n,p}^Q(\lambda_2^*, \ldots, \lambda_p^*) = C_{n-1,p-1,\beta} \prod_{2 \le i < j \le p} |\lambda_i^* - \lambda_j^*|^\beta \prod_{i=2}^{p} (\lambda_i^*)^{\beta(n-p+1)/2-1} \times e^{-(n-1)\sum_{i=2}^p \lambda_i^*/2}.$$

This implies that the density function of $(\lambda_2, \ldots, \lambda_p)$ under Q is

$$\begin{aligned} & f_{n,p}^Q(\lambda_2, \ldots, \lambda_p) \\ = & \left(\frac{n}{n-1}\right)^{\beta(n-1)(p-1)/2} C_{n-1,p-1,\beta} \prod_{2 \le i < j \le p} |\lambda_i - \lambda_j|^\beta \times \prod_{i=2}^{p} \lambda_i^{\beta(n-p+1)/2-1} \times e^{-n\sum_{i=2}^p \lambda_i/2}. \end{aligned} \tag{4.11}$$

Therefore dQ/dP takes the form

$$\begin{aligned} & \frac{f_{n,p}^Q(\lambda_2, \ldots, \lambda_p) \times nre^{-nr(\lambda_1 - \tilde{x} \vee \lambda_2)} \times \mathbf{1}(\lambda_1 > \tilde{x} \vee \lambda_2)}{f_{n,p}(\lambda_1, \ldots, \lambda_p)} \\ = & \frac{\left(\frac{n}{n-1}\right)^{\beta(n-1)(p-1)/2} C_{n-1,p-1,\beta}\, nre^{-nr(\lambda_1 - \tilde{x} \vee \lambda_2)} \times \mathbf{1}(\lambda_1 > \tilde{x} \vee \lambda_2)}{C_{n,p,\beta} \prod_{i=2}^p (\lambda_1 - \lambda_i) \times \lambda_1^{\beta(n-p+1)/2-1} \times e^{-n\lambda_1/2}}. \end{aligned}$$

The corresponding importance sampling estimate is given by

$$L_{n,p}(x) = \mathbf{1}(U_{n,p} > x)\, dP/dQ, \tag{4.12}$$

where $U_{n,p}$ is calculated with the sampled $\lambda_1, \ldots, \lambda_p$ based on Eq. (4.1).

We claim that for the proposed Algorithm IV.1, with the choice of r specified in (4.9), the importance sampling estimator $L_{n,p}(x)$ is asymptotically efficient in estimating the target tail probability. This result is formally stated below and proved in Section 4.5.

Theorem 4.2.1. *When $p/n \to \gamma \in \mathbb{R}$, the estimator $L_{n,p}(x)$ in (4.12) is an asymptotically efficient estimator of $\alpha_{n,p}(x)$ for $x > (\sqrt{\gamma}+1)^2$.*

Remark IV.1. *Our discussion regarding asymptotic efficiency focuses on the case of estimating rare-event tail probability* $\alpha_{n,p}(x)$, *i.e., when* $\{U_{n,p} > x\}$ *corresponds to a rare event. When* $x \leq (\sqrt{\gamma}+1)^2$, $\{U_{n,p} > x\}$ *is not rare, and we can still apply the importance sampling algorithm with a reasonable positive* r *value as the exponential distribution's rate. However, the theoretical properties of the importance sampling estimator must then be studied under a different framework; this issue is not pursued here.*

Remark IV.2. *We explain the Marchenko–Pastur form of (4.10). When the entries of* $\mathbf{X}$ *have mean 0 and variance 1 (*$\beta = 1$ *and* 2*), the Marchenko–Pastur law for the eigenvalues of* $\mathbf{X}^H\mathbf{X}/n$ *takes the standard form*

$$f(d\bar{s}) = (2\pi\gamma\bar{s})^{-1}\sqrt{(\bar{s}_+ - \bar{s})(\bar{s} - \bar{s}_-)}\,\mathbf{1}(\bar{s} \in [\bar{s}_-, \bar{s}_+])d\bar{s} \tag{4.13}$$

with $\bar{s}_- = (1-\sqrt{\gamma})^2$ *and* $\bar{s}_+ = (1+\sqrt{\gamma})^2$ *(see, e.g., Paul and Aue, 2014, Theorem 3.2). For the setting considered of this chapter, the real case (*$\beta = 1$*) has* $\sigma^2 = 1$, *so* (4.10) *and* (4.13) *are consistent. In contrast, the complex case (*$\beta = 2$*) has* $\sigma^2 = 2$ *and therefore* (4.10) *and* (4.13) *are different up to a factor of* $\beta = 2$. *Specifically, let* $(\bar{\lambda}_1, \ldots, \bar{\lambda}_p)$ *and* $(\lambda_1, \ldots, \lambda_p)$ *be eigenvalues of* $\mathbf{X}^H\mathbf{X}/n$ *when* $\mathbf{X}$ *has i.i.d. entries of* $\mathcal{CN}(0,1)$ *and* $\mathcal{CN}(0,2)$, *respectively. Then we know that* $(\lambda_1, \ldots, \lambda_p) \sim 2(\bar{\lambda}_1, \ldots, \bar{\lambda}_p)$ *and* (4.13) *implies the empirical distribution in (4.10).*

Remark IV.3. *We discuss the differences between the proposed method and the method in Jiang et al. (2017) on the largest eigenvalue, which also employs an importance sampling technique. First, the two methods have different targets, i.e.,* $\Pr(\lambda_1 > x)$ *in Jiang et al. (2017) and* $\Pr(U_{n,p} > x)$ *here, and therefore use different changes of measure to construct efficient importance sampling estimators. As discussed in Section 4.2, in order to achieve asymptotic efficiency, the change of measures should approximate the target conditional distribution measures, i.e.,* $\Pr(\,\cdot \mid \lambda_1 > x)$ *in*

Jiang et al. (2017) and $\Pr(\,\cdot \mid U_{n,p} > x)$ *in this chapter. Due to the difference between the two conditional distributions, different changes of measure are constructed in the two methods. Specifically, Jiang et al. (2017) sample the largest eigenvalue* λ_1 *from a truncated exponential distribution depending on the second largest eigenvalue* λ_2*, while the present work samples* λ_1 *from an exponential distribution depending on eigenvalues* $\lambda_2, \ldots, \lambda_p$*. Second, the proof techniques of the main asymptotic results in the two papers are also different. In particular, to show the asymptotic efficiency of the importance sampling estimators as defined in* (4.5)*, we need to derive asymptotic approximations for both the rare-event probability* $\alpha_{n,p}(x)$ *and the second moments of the importance sampling estimator* $\mathrm{E}_Q\{L_{n,p}^2(x)\}$*. Even though the largest eigenvalue* λ_1 *and the ratio statistic* $U_{n,p}$ *have similar large deviation approximation results for their tail probabilities, the asymptotic approximations for the second moments of the importance sampling estimators are different due to the differences between the considered changes of measure as well as the effect of the trace term in* $U_{n,p}$*. Please refer to the proof for more details.*

Remark IV.4. *The method and the theoretical results can be easily extended from the case* $p \leq n$ *to the case* $p \geq n$ *by switching the labels of* n *and* p *and changing* γ *to* $1/\gamma$ *correspondingly. Note that when* $p \geq n$*, the eigenvalues of* $\mathbf{X}^H\mathbf{X}/n$ *and* $\mathbf{X}\mathbf{X}^H/p$ *give the same test statistic* $U_{n,p}$ *as defined in* (4.1)*, which is because* $\mathbf{X}^H\mathbf{X}$ *and* $\mathbf{X}\mathbf{X}^H$ *have the same set of nonzero eigenvalues and* $U_{n,p}$ *is scale invariant. By symmetry, when* $p \geq n$*, the joint density function of the eigenvalues of* $\mathbf{X}\mathbf{X}^H/p$ *have the same form as* (4.4)*, except that the labels of* n *and* p *are switched. Therefore, the cases when* $p \leq n$ *and* $p \geq n$ *are equivalent up to the label switching. Note that after* p/n *is changed to* n/p*,* γ *becomes* $1/\gamma$ *correspondingly.*

4.3 Simulation Studies

We conducted simulation studies to evaluate the performance of our algorithm. We first took combinations $(n, p) \in \{(100, 10), (100, 20), (500, 20), (1000, 50)\}$, and $\beta = 1$ and 2, respectively. Then we compared our algorithm with other methods and present the results in Table 4.1 and 4.2.

For the proposed importance sampling estimator, we repeated $N_{IS} = 10^4$ times and show the estimated probabilities ("EST_{IS}" column) along with the estimated standard deviations of $L_{n,p}$, i.e., $\sqrt{Var^Q(L_{n,p})}$ ("SD_{IS}" column). The ratios between estimated standard deviations and estimates ("SD_{IS}/EST_{IS}" column) reflect the efficiency of the algorithms. Note that with $N_{IS} = 10^4$ replications, the standard error of the estimate is $SD_{IS}/\sqrt{N_{IS}} = SD_{IS}/100$.

In addition, three alternative methods were considered, namely the direct Monte Carlo, the Tracy–Widom distribution approximation, and the corrected Tracy–Widom approximation (Nadler, 2011). We computed direct Monte Carlo estimates ("EST_{DMC}" column) with $N_{DMC} = 10^6$ independent replications. We present the standard deviation of direct Monte Carlo estimates ("SD_{DMC}" column) and the ratios between estimated standard deviations and estimates ("SD_{DMC}/EST_{DMC}"). In addition, we used the approximation of Tracy–Widom distribution ("TW" column) specified in Eq. (4.2). The $TW(x)$ is computed from the `RMTstat` package in R. Furthermore, following Nadler (2011), we computed the Tracy–Widom approximation with correction term ("$c.TW$" column), viz.,

$$\Pr\left(\frac{U - \mu_{n,p}}{\sigma_{n,p}} > x\right) \approx 1 - \mathcal{TW}_\beta(x) + \frac{1}{2}\left(\frac{2}{np}\right)\left(\frac{\mu_{n,p}}{\sigma_{n,p}}\right)^2 \mathcal{TW}_\beta''(x), \qquad (4.14)$$

where $\mathcal{TW}''(x)$ is computed numerically via a standard central differencing scheme with $\Delta x = 10^{-3}$. When $\beta = 1$, μ and σ are chosen according to Eq. (4.3). When $\beta = 2$, μ and σ are chosen according to Karoui (2006).

We can see from Tables 4.1 and 4.2 that the Tracy–Widom distribution ("TW" column) significantly overestimates the tail probabilities for all considered settings and the finding is consistent with that in Nadler (2011). Furthermore, the corrected Tracy–Widom approximation ("$c.TW$" column) underestimates the tail probability $\alpha_{n,p}(x)$ and goes to a negative number as $\alpha_{n,p}(x)$ becomes small.

Since the proposed importance sampling and the direct Monte Carlo method are both unbiased estimators, next we compare their computational efficiency. As discussed in Section 4.2, for the average estimator $\bar{L}_{n,p}(x) = \{L_{n,p}^{(1)}(x) + \cdots + L_{n,p}^{(N)}(x)\}/N$, "$SD_{IS}/EST_{IS}$" and "$SD_{DMC}/EST_{DMC}$" can be used as a measure of the computational efficiency in terms of iteration numbers. From the results in Tables 4.1 and 4.2, as $\alpha_{n,p}(x)$ decreases, "SD_{DMC}/EST_{DMC}" grows quickly and even becomes not available. In contrast, "SD_{IS}/EST_{IS}" increases slowly and is generally smaller than "SD_{DMC}/EST_{DMC}", showing that the proposed importance sampling is more efficient than the direct Monte Carlo method.

As a further illustration, we compared the iteration numbers N_{IS} and N_{DMC} that would be needed to achieve the same level of relative standard errors of the estimators. Specifically, in order to have the same ratios of the standard errors to the estimates, i.e., $SE_{IS}/EST_{IS} = (SD_{IS}/\sqrt{N_{IS}})/EST_{IS}$ and $SE_{DMC}/EST_{DMC} = (SD_{DMC}/\sqrt{N_{DMC}})/EST_{DMC}$, obtained under the importance sampling and direct direct Monte Carlo, respectively, we need

$$\frac{N_{DMC}}{N_{IS}} = \frac{(SD_{DMC}/EST_{DMC})^2}{(SD_{IS}/EST_{IS})^2}. \tag{4.15}$$

Based on the above equation, the simulation results show that to have a similar standard error obtained under the importance sampling, the direct Monte Carlo method needs more iterations as $\alpha_{n,p}(x)$ goes small. For example, from Table 1, when $n = 100$, $p = 10$ and $x = 2.1$, we need N_{DMC} to be approximately 4.3×10^2 times larger than

Table 4.1: Results of estimating tail probabilities of $U_{n,p}$ in (4.1) for the real Wishart matrix ($\beta = 1$).

(a) $n = 100$, $p = 10$

x	EST_{IS}	SD_{IS}	$\text{SD}_{\text{IS}}/\text{EST}_{\text{IS}}$	EST_{DMC}	SD_{DMC}	$\text{SD}_{\text{DMC}}/\text{EST}_{\text{DMC}}$	c.TW	TW
1.80	2.44e-2	1.25e-1	5.14	2.46e-2	1.55e-1	6.30	2.58e-2	5.07e-2
1.95	1.02e-3	5.00e-3	4.89	1.08e-3	3.28e-2	30.46	3.90e-4	4.37e-3
1.98	5.32e-4	3.55e-3	6.66	5.57e-4	2.36e-2	42.36	4.96e-6	2.48e-3
2.10	2.43e-5	2.48e-4	10.22	2.20e-5	4.69e-3	213.20	–7.46e-5	2.07e-4
2.30	5.25e-8	7.72e-7	14.71	0	0	NaN	0	0

(b) $n = 100$, $p = 20$

x	EST_{IS}	SD_{IS}	$\text{SD}_{\text{IS}}/\text{EST}_{\text{IS}}$	EST_{DMC}	SD_{DMC}	$\text{SD}_{\text{DMC}}/\text{EST}_{\text{DMC}}$	c.TW	TW
2.10	9.14e-2	3.73e-1	4.09	8.99e-2	2.86e-1	3.18	9.29e-2	1.21e-1
2.30	2.86e-3	2.04e-2	7.13	2.71e-3	5.20e-2	19.19	2.31e-3	6.09e-3
2.40	3.44e-4	2.60e-3	7.54	3.11e-4	1.76e-2	56.70	1.54e-4	9.07e-4
2.50	2.89e-5	2.01e-4	6.95	2.60e-5	5.10e-3	196.11	–6.13e-6	1.05e-4
2.70	1.50e-7	1.78e-6	11.85	0	0	NaN	0	0

(c) $n = 500$, $p = 20$

x	EST_{IS}	SD_{IS}	$\text{SD}_{\text{IS}}/\text{EST}_{\text{IS}}$	EST_{DMC}	SD_{DMC}	$\text{SD}_{\text{DMC}}/\text{EST}_{\text{DMC}}$	c.TW	TW
1.46	4.64e-2	2.21e-1	4.76	4.68e-2	2.11e-1	4.51	4.87e-2	6.51e-2
1.51	3.98e-3	2.16e-2	5.43	3.70e-3	6.07e-2	16.40	3.70e-3	7.03e-3
1.56	1.57e-4	7.13e-4	4.54	1.55e-4	1.24e-2	80.32	1.28e-4	4.40e-4
1.62	2.14e-6	1.49e-5	6.97	3.00e-6	1.73e-3	577.35	–1.87e-6	6.71e-6
1.70	2.43e-9	2.72e-8	11.20	0	0	NaN	0	0

(d) $n = 1000$, $p = 50$

x	EST_{IS}	SD_{IS}	$\text{SD}_{\text{IS}}/\text{EST}_{\text{IS}}$	EST_{DMC}	SD_{DMC}	$\text{SD}_{\text{DMC}}/\text{EST}_{\text{DMC}}$	c.TW	TW
1.52	2.75e-2	1.29e-1	4.70	2.90e-2	1.68e-1	5.78	2.96e-2	3.59e-2
1.55	2.51e-3	1.16e-2	4.63	2.57e-3	5.06e-2	19.71	2.53e-3	7.98e-4
1.60	1.41e-5	5.25e-5	3.72	2.20e-5	4.69e-3	213.20	1.15e-5	3.25e-5
1.62	1.40e-6	8.70e-6	6.21	2.00e-6	1.41e-3	707.11	–7.93e-7	6.71e-6
1.66	7.49e-9	3.69e-8	4.93	0	0	NaN	0	0

Table 4.2: Results of estimating tail probabilities of $U_{n,p}$ in (4.1) for the complex Wishart matrix ($\beta = 2$).

(a) $n = 100$, $p = 10$

x	EST_{IS}	SD_{IS}	$\text{SD}_{\text{IS}}/\text{EST}_{\text{IS}}$	EST_{DMC}	SD_{DMC}	$\text{SD}_{\text{DMC}}/\text{EST}_{\text{DMC}}$	c.TW	TW
1.77	3.72e-3	3.34e-2	8.98	3.79e-3	6.15e-2	16.21	2.20e-3	1.26e-2
1.81	9.21e-4	1.32e-2	14.34	8.97e-4	2.99e-2	33.37	-1.36e-4	4.42e-3
1.91	1.89e-5	3.28e-4	17.37	1.70e-5	4.12e-3	242.53	-1.22e-4	2.11e-4
1.93	6.68e-6	8.44e-5	12.64	4.00e-6	2.00e-3	500	-7.44e-5	1.07e-4
1.99	2.98e-7	4.25e-6	14.27	0	0	NaN	-1.29e-5	1.24e-5

(b) $n = 100$, $p = 20$

x	EST_{IS}	SD_{IS}	$\text{SD}_{\text{IS}}/\text{EST}_{\text{IS}}$	EST_{DMC}	SD_{DMC}	$\text{SD}_{\text{DMC}}/\text{EST}_{\text{DMC}}$	c.TW	TW
2.10	1.20e-2	7.99e-2	6.68	1.45e-2	1.20e-1	8.23	1.41e-2	2.70e-2
2.18	1.04e-3	7.59e-3	7.28	1.34e-3	3.66e-2	27.29	8.64e-4	3.65e-3
2.30	2.18e-5	3.47e-4	15.94	2.30e-5	4.80e-3	208.51	-2.06e-5	8.86e-5
2.38	6.73e-7	1.94e-5	28.86	1.00e-6	1.00e-3	1000	-2.70e-6	4.83e-6
2.46	1.63e-8	2.83e-7	17.36	0	0	NaN	-1.73e-7	1.93e-7

(c) $n = 500$, $p = 20$

x	EST_{IS}	SD_{IS}	$\text{SD}_{\text{IS}}/\text{EST}_{\text{IS}}$	EST_{DMC}	SD_{DMC}	$\text{SD}_{\text{DMC}}/\text{EST}_{\text{DMC}}$	c.TW	TW
1.45	8.04e-3	5.49e-4	6.84	8.98e-3	9.43e-2	10.51	8.95e-3	1.58e-2
1.48	6.56e-4	8.02e-3	12.22	6.49e-4	2.55e-2	39.24	5.07e-4	1.59e-3
1.50	8.77e-5	1.16e-3	13.18	8.60e-5	9.27e-3	107.83	3.88e-5	2.70e-4
1.525	5.05e-6	5.37e-5	10.63	8.00e-6	2.83e-3	353.55	-1.87e-6	2.28e-5
1.55	1.85e-7	1.71e-6	9.28	0	0	NaN	-4.66e-7	1.49e-6

(d) $n = 1000$, $p = 50$

x	EST_{IS}	SD_{IS}	$\text{SD}_{\text{IS}}/\text{EST}_{\text{IS}}$	EST_{DMC}	SD_{DMC}	$\text{SD}_{\text{DMC}}/\text{EST}_{\text{DMC}}$	c.TW	TW
1.51	5.85e-3	6.67e-2	11.39	5.20e-3	7.19e-2	13.83	5.31e-3	7.46e-3
1.53	2.65e-4	1.96e-3	7.39	3.04e-4	1.74e-2	57.35	2.98e-4	5.32e-4
1.56	1.72e-6	1.84e-5	10.72	0	0	NaN	1.33e-6	4.20e-6
1.58	3.15e-8	2.86e-7	9.10	0	0	NaN	1.46e-8	9.85e-8
1.60	4.24e-10	3.80e-9	8.97	0	0	NaN	-6.21e-11	1.56e-9

N_{IS}; when $n = 1000$, $p = 50$ and $x = 1.62$, we need N_{DMC} to be about 1.3×10^4 times larger.

Besides the iteration numbers, we compared the average time cost of each iteration under the importance sampling and the direct Monte Carlo method, respectively. For the direct Monte Carlo, two methods were considered in computing the eigenvalues. The first method directly computes the test statistic $U_{n,p}$ using the eigen-decomposition of a randomly sampled Wishart matrix. The second method computes the eigenvalues from the tridiagonal representation form as in Step 1 of Algorithm 1. We ran 10^4 iterations for all the methods and report the average time of one iteration in Table 4.3, where the first method of the direct Monte Carlo is denoted as T_{DMC_1}, the second method is denoted as T_{DMC_2}, and the importance sampling method is denoted as T_{IS}. The simulation results show that T_{DMC_1} has the highest time cost per iteration, while T_{DMC_2} and T_{IS} are similar.

We further explain the simulation results from the perspective of algorithm complexity. For each iteration, the first direct Monte Carlo method samples a $p \times p$ Wishart matrix and performs its eigen-decomposition, whose cost is typically of the order of $O(p^3)$. The second direct Monte Carlo method and the importance sampling only need to sample $O(p)$ number of chi-square random variables and then decompose a symmetric tridiagonal matrix, at a cost of $O(p^2)$ per iteration Demmel (1997). Although the importance sampling also samples from an exponential distribution in Step 2, the distribution parameters can be calculated in advance and it does not affect the overall complexity much. Therefore, the time complexity of the algorithm T_{DMC_1} is higher while T_{DMC_2} and T_{IS} are similar per iteration. Together with the result in (4.15), we can see that the importance sampling is more efficient than the direct Monte Carlo method in terms of both the iteration number and the overall time cost.

To further check the influence of replication number N_{IS} of the importance sampling algorithm, we focus on the case $(n, p) = (100, 10)$ and compare the performance of different N_{IS}s. In order to obtain accurate reference values of the tail probabilities, we used direct Monte Carlo with repeating time $N_{DMC} = 10^8$ to estimate multiple tail probabilities $\alpha_{n,p}(x)$s ranging from 10^{-2} to 10^{-6} under $\beta = 1$ and $\beta = 2$, respectively. Then we estimated the corresponding $\alpha_{n,p}(x)$s using our algorithm with $N_{IS} = 10^4, 10^5, 10^6$, respectively.

The results are presented in Figure IV.1, where the x-axis represents the reference values $\log_{10}(EST_{DMC})$. The line "*DMC with error bar*" represents the (estimated) pointwise 95% confidence intervals, viz.,

$$[\log_{10}(EST_{DMC} - 2 \times SD_{DMC}/\sqrt{N_{DMC}}), \log_{10}(EST_{DMC} + 2 \times SD_{DMC}/\sqrt{N_{DMC}})].$$

Similarly, the line "*Importance Sampling with error bar*" represents the importance sampling estimates and pointwise 95% confidence intervals, viz.,

$$[\log_{10}(EST_{IS} - 2 \times SD_{IS}/\sqrt{N_{IS}}), \log_{10}(EST_{IS} + 2 \times SD_{IS}/\sqrt{N_{IS}})].$$

One can surmise from the figures that the proposed algorithm gives reliable estimates of probabilities as small as 10^{-6} with $N_{IS} = 10^4$, which is more efficient than directed Monte Carlo and more accurate than Tracy–Widom approximations. Furthermore, Figure IV.1 shows that the algorithm improves as the number of iterations increases. We also plot the Tracy–Widom approximations in (4.2) and (4.14) in Figure IV.1 for comparison.

Figure IV.1 shows that without correction, the Tracy–Widom distribution in (4.2) is not accurate and overestimates the probabilies. The correction term in (4.14) improves the approximation when the probability is larger than the scale of about 10^{-2}, which is consistent with the result in Nadler (2011). But when the probability gets smaller, the corrected approximation has larger deviation from true values (on the $\log_{10}$ scale) and even becomes negative. Note that since we cannot plot the

$\log_{10}$ of negative numbers in the figures, the lines of the corrected Tracy–Widom approximations appear to be shorter. These results validate the results in Table 4.1 and 4.2.

Table 4.3: Estimated computation time of three sampling methods.

(a) $\beta = 1$

n	p	x	T_{DMC_1}	T_{DMC_2}	T_{IS}
100	10	1.95	1.28e-03	7.26e-04	8.89e-05
100	10	1.98	1.15e-03	9.23e-05	8.51e-05
100	20	2.3	1.58e-03	7.35e-05	6.84e-05
100	20	2.4	1.65e-03	1.79e-04	6.33e-05
500	20	1.51	1.27e-03	9.87e-05	9.32e-05
500	20	1.56	1.67e-03	7.39e-05	8.82e-05
1000	50	1.55	3.19e-03	1.05e-04	1.56e-04
1000	50	1.6	3.12e-03	9.76e-05	1.34e-04

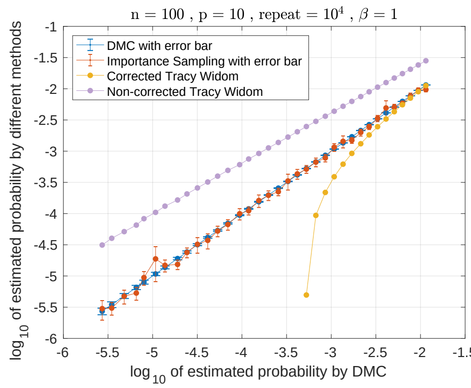
(b) $\beta = 2$

n	p	x	T_{DMC_1}	T_{DMC_2}	T_{IS}
100	10	1.77	1.87e-03	1.75e-04	6.08e-05
100	10	1.81	1.85e-03	5.47e-05	5.90e-05
100	20	2.18	2.86e-03	8.37e-05	1.20e-04
100	20	2.3	2.69e-03	1.11e-04	6.69e-05
500	20	1.45	2.79e-03	8.46e-05	7.01e-05
500	20	1.48	3.53e-03	7.24e-05	8.90e-05
1000	50	1.53	8.65e-03	9.03e-05	1.53e-04
1000	50	1.56	8.35e-03	9.61e-05	1.49e-04

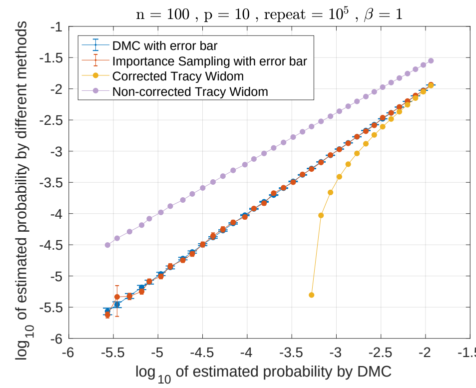
4.4 Discussion

This chapter proposes an asymptotically efficient Monte Carlo method to estimate the tail probabilities of the ratio of the largest eigenvalue to the trace of the Wishart matrix. Theoretically, we prove that the importance sampling estimator is asymptotically efficient. Numerically, we conduct extensive studies to evaluate the performance of the proposed algorithm compared with other methods in terms of estimation accuracy and computational cost in estimating the tail probabilities.

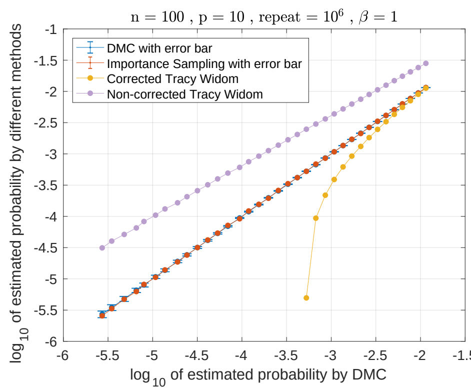
The method can be adapted to estimate tail probabilities of the ratio of the sum



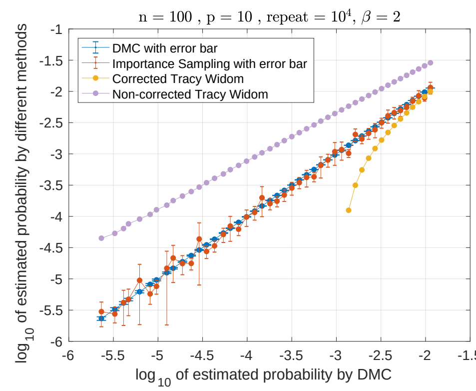
(a)



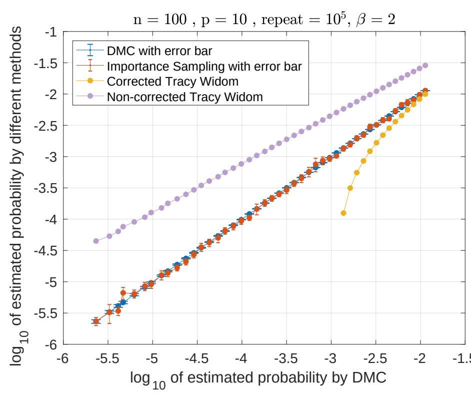
(b)



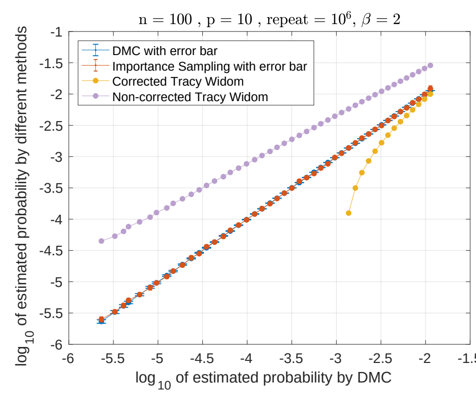
(c)



(d)



(e)



(f)

Figure IV.1: $\log_{10}$ of estimated probabilities by four compared methods versus $\log_{10}$ of estimated probabilities by DMC when $n = 100$ and $p = 10$.

Table 4.4: Results of estimating tail probabilities of $U_{n,p}^k$ in (4.16) for the real Wishart matrix ($\beta = 1$).

(a) $n = 100$, $p = 50$, $k = 2$

x	$\mathrm{EST_{IS}}$	$\mathrm{SD_{IS}}$	$\mathrm{SD_{IS}/EST_{IS}}$	$\mathrm{EST_{DMC}}$	$\mathrm{SD_{DMC}}$	$\mathrm{SD_{DMC}/EST_{DMC}}$
5.9	1.14e-03	6.70e-03	5.89	1.55e-03	3.93e-02	25.41
6.0	3.22e-04	3.80e-03	11.78	2.98e-04	1.73e-02	57.92
6.1	5.68e-05	9.37e-04	16.49	5.50e-05	7.42e-03	134.84
6.4	1.09e-07	3.21e-06	29.50	0	0	NaN

(b) $n = 100$, $p = 50$, $k = 3$

x	$\mathrm{EST_{IS}}$	$\mathrm{SD_{IS}}$	$\mathrm{SD_{IS}/EST_{IS}}$	$\mathrm{EST_{DMC}}$	$\mathrm{SD_{DMC}}$	$\mathrm{SD_{DMC}/EST_{DMC}}$
8.4	1.56e-03	1.77e-02	11.36	1.55e-03	3.93e-02	25.41
8.5	4.22e-04	5.48e-03	12.98	4.04e-04	2.01e-02	49.74
8.7	1.46e-05	2.99e-04	20.44	1.80e-05	4.24e-03	235.70
8.9	7.26e-07	2.53e-05	34.83	0	0	NaN

(c) $n = 100$, $p = 50$, $k = 4$

x	$\mathrm{EST_{IS}}$	$\mathrm{SD_{IS}}$	$\mathrm{SD_{IS}/EST_{IS}}$	$\mathrm{EST_{DMC}}$	$\mathrm{SD_{DMC}}$	$\mathrm{SD_{DMC}/EST_{DMC}}$
10.6	7.60e-03	5.65e-02	7.43	8.01e-03	8.91e-02	11.13
10.8	6.58e-04	6.63e-03	10.08	8.44e-04	2.90e-02	34.41
11.0	5.49e-05	1.47e-03	26.73	6.40e-05	8.00e-03	125.00
11.3	1.70e-07	5.56e-06	32.77	0	0	NaN

Table 4.5: Results of estimating tail probabilities of $U^k_{n,p}$ in (4.16) for the complex Wishart matrix ($\beta = 2$).

(a) $n = 100$, $p = 50$, $k = 2$

x	$\mathrm{EST}_{\mathrm{IS}}$	$\mathrm{SD}_{\mathrm{IS}}$	$\mathrm{SD}_{\mathrm{IS}}/\mathrm{EST}_{\mathrm{IS}}$	$\mathrm{EST}_{\mathrm{DMC}}$	$\mathrm{SD}_{\mathrm{DMC}}$	$\mathrm{SD}_{\mathrm{DMC}}/\mathrm{EST}_{\mathrm{DMC}}$
5.6	4.67e-03	3.52e-02	7.54	5.03e-03	7.07e-02	14.07
5.7	5.08e-04	5.95e-03	11.72	4.98e-04	2.23e-02	44.80
5.8	4.75e-05	9.55e-04	20.12	3.80e-05	6.16e-03	162.23
6.0	7.71e-08	2.48e-06	32.18	0	0	NaN

(b) $n = 100$, $p = 50$, $k = 3$

x	$\mathrm{EST}_{\mathrm{IS}}$	$\mathrm{SD}_{\mathrm{IS}}$	$\mathrm{SD}_{\mathrm{IS}}/\mathrm{EST}_{\mathrm{IS}}$	$\mathrm{EST}_{\mathrm{DMC}}$	$\mathrm{SD}_{\mathrm{DMC}}$	$\mathrm{SD}_{\mathrm{DMC}}/\mathrm{EST}_{\mathrm{DMC}}$
8.1	1.78e-03	2.08e-02	11.67	2.16e-03	4.64e-02	21.50
8.2	3.67e-04	8.31e-03	22.67	2.90e-04	1.70e-02	58.71
8.3	1.87e-05	3.73e-04	19.96	2.50e-05	5.00e-03	200.00
8.5	1.50e-07	6.90e-06	45.95	0	0	NaN

(c) $n = 100$, $p = 50$, $k = 4$

x	$\mathrm{EST}_{\mathrm{IS}}$	$\mathrm{SD}_{\mathrm{IS}}$	$\mathrm{SD}_{\mathrm{IS}}/\mathrm{EST}_{\mathrm{IS}}$	$\mathrm{EST}_{\mathrm{DMC}}$	$\mathrm{SD}_{\mathrm{DMC}}$	$\mathrm{SD}_{\mathrm{DMC}}/\mathrm{EST}_{\mathrm{DMC}}$
10.4	2.49e-03	4.78e-02	19.18	2.73e-03	5.22e-02	19.12
10.5	4.27e-04	6.15e-03	14.40	4.42e-04	2.10e-02	47.55
10.6	5.47e-05	1.86e-03	34.04	6.90e-05	8.31e-03	120.38
10.8	3.17e-07	1.23e-05	38.96	0	0	NaN

of the first k largest eigenvalues to the trace of the Wishart matrix, which is defined as

$$U_{n,p}^k = \frac{\lambda_1 + \cdots + \lambda_k}{(\lambda_1 + \cdots + \lambda_p)/\min(p,n)}, \tag{4.16}$$

where k is a fixed positive integer. The algorithm is as follows. First sample $\lambda_2, \ldots, \lambda_p$ from $\mathbf{L}_{n-1,p-1,\beta}/n$ using the same method as in Algorithm IV.1. Second, conditioning on $\lambda_2, \ldots, \lambda_p$, sample λ_1 from a truncated exponential distribution of the same form as (4.8), but redefine

$$\tilde{x} = \frac{1}{p-x}\left(x\sum_{i=2}^{p}\lambda_i - p\sum_{i=2}^{k}\lambda_i\right)$$

and choose r to be a small constant that depends on the large deviation result of the largest k eigenvalues.

We conducted a numerical study to show the validity and efficiency of the proposed method in estimating the tail probabilities of $U_{n,p}^k$. Following the design in Section 4.3, the sampling was repeated 10^4 times for the importance sampling method and 10^6 times for the direct Monte Carlo method. The constant k was chosen to be $2, 3, 4$, and we took $n = 100$, $p = 50$, and $r = 1/10$. Tables 4.4 and 4.5 summarize the results of $\beta = 1$ and $\beta = 2$, which show similar patterns as Tables 4.1 and 4.2. When the tail probability becomes smaller, SD_{IS}/EST_{IS} is smaller than SD_{DMC}/EST_{DMC}, which indicates that the importance sampling is more efficient than the direct Monte Carlo method in estimating the tail probabilities, as discussed in Section 4.3. It would be interesting to study the asymptotic property of this algorithm on estimating the tail probability of $U_{n,p}^k$. However, this would require the development of asymptotic theory on the tail probabilities of the first k largest eigenvalues, which is beyond the scope of this study. We leave it for future work.

4.5 Proofs

This section provides the proof of Theorem 4.2.1 on the estimator's asymptotic efficiency. We focus on the case when $p \le n$ and $p/n \to \gamma \in (0, 1]$. For the case of $p \ge n$ and $p/n \to \gamma \in [1, \infty)$, the proof follows from the same argument by switching the labels of n and p, as shown in Remark IV.4.

Recall the definition of Q, $L_{n,p} = \mathbf{1}(U_{n,p} > x)dP/dQ$ and $\alpha_{n,p}(x) = \Pr(U_{n,p} > x)$. To prove the asymptotic efficiency defined in (4.5), we need only show that $\liminf_{n\to\infty} \ln \mathrm{E}_Q(L_{n,p}^2)/\{2\ln \alpha_{n,p}(x)\} \ge 1$ since $\mathrm{E}_Q(L_{n,p}^2) \le \mathrm{var}_Q(L_{n,p}^2)$. We give an outline of the proof first.

Step 1. We give the asymptotic approximation of $\lim_{n\to\infty} n^{-1} \ln \alpha_{n,p}(x) = -\gamma I_\beta(\beta x)$, where $I_\beta(\beta x)$ is the large deviation rate function.

Step 2. By the result in Step 1, we only need to prove that

$$\liminf_{n\to\infty} \frac{\ln \mathrm{E}_Q(L_{n,p}^2)}{2\ln \alpha_{n,p}(x)} = \liminf_{n\to\infty} \frac{\ln \mathrm{E}_Q(L_{n,p}^2)}{-2\gamma I_\beta(\beta x)} \ge 1.$$

This is established using the upper bound $I_1 + I_2 + I_3$ of $\mathrm{E}_Q(L_{n,p}^2)$ in (4.21) together with the limiting properties of I_1, I_2, and I_3 in (4.22), (4.23), and (4.24), respectively.

The details of Steps 1 and 2 are given below.

Step 1. We first obtain the large deviation rate function for $U_{n,p}$, which gives an approximation to $n^{-1} \ln \alpha_{n,p}(x)$ as in Anderson et al. (2010). From the argument in Bianchi et al. (2011), the large deviation of $U_{n,p}$ has a similar rate function as λ_1. The explicit form of the large deviation rate function of λ_1 can be obtained from Theorem 2.6.6 in Anderson et al. (2010). In particular, denote $(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_p)$ to be the unordered eigenvalues of $\mathbf{X}^H\mathbf{X}/n$; then from (4.4), $(\tilde\lambda_1,\ldots,\tilde\lambda_p)$ has joint density function

$$\begin{aligned} f_{n,p,\beta}(\tilde\lambda_1,\ldots,\tilde\lambda_p) &= \frac{1}{p!}C_{n,p,\beta}\prod_{1\le i<j\le p}|\tilde\lambda_i-\tilde\lambda_j|^\beta \times \prod_{i=1}^{p}\tilde\lambda_i^{\beta(n-p+1)/2-1}\times e^{-n\sum_{i=1}^p\tilde\lambda_i/2} \\ &= (Z_{V,\beta}^p)^{-1}|\Delta_p(\tilde\lambda)|^\beta e^{-p\sum_{i=1}^p V(\tilde\lambda_i)}, \end{aligned}$$

where the last line follows the notation of (2.6.1) in Anderson et al. (2010) with $\Delta_p(\tilde\lambda)=\prod_{1\le i<j\le p}(\tilde\lambda_i-\tilde\lambda_j)$, $Z_{V,\beta}^p = p!C_{n,p,\beta}^{-1}$ and

$$V(x)=\frac{n}{2p}x-\frac{\beta(n-p+1)-2}{2p}\ln x \;\sim\; \frac{1}{2}\left\{\frac{x}{\gamma}-\beta\left(\frac{1}{\gamma}-1\right)\ln x\right\}.$$

The notation "$a_n\sim b_n$" means $a_n=\{1+o(1)\}b_n$. Following the definition in (2.6.3) of Anderson et al. (2010), we further define

$$\begin{aligned} Z_{pV/(p-1),\beta}^{p-1} &= \int_{\mathbb{R}}\cdots\int_{\mathbb{R}}|\Delta_{p-1}(\tilde\lambda)|^\beta e^{-(p-1)\sum_{i=1}^{p-1}\left\{pV(\tilde\lambda_i)/(p-1)\right\}}\prod_{i=1}^{p-1}d\tilde\lambda_i \\ &= \int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\prod_{1\le i<j\le(p-1)}|\tilde\lambda_i-\tilde\lambda_j|^\beta\prod_{i=1}^{p-1}\tilde\lambda_i^{\beta(n-p+1)/2-1}\times e^{-n\sum_{i=1}^{p-1}\tilde\lambda_i/2}\prod_{i=1}^{p-1}d\tilde\lambda_i. \end{aligned}$$

Then, the density function (4.11) implies that the normalization constant $Z_{pV/(p-1),\beta}^{p-1}$ equals

$$Z_{pV/(p-1),\beta}^{p-1} = \left\{\frac{1}{(p-1)!}\left(\frac{n}{n-1}\right)^{\beta(n-1)(p-1)/2}C_{n-1,p-1,\beta}\right\}^{-1}.$$

With the above notation, Theorem 2.6.6 in Anderson et al. (2010) states that the large deviation approximation of $\lambda_1=\max(\tilde\lambda_1,\ldots,\tilde\lambda_p)$ has speed p and good rate function, viz.

$$I_\beta(s)=\begin{cases} -\beta\int_{\mathbb{R}}\ln|s-t|\sigma_\beta(dt)+V(s)+\alpha_{V,\beta} & \text{if } s\ge s^*, \\ \infty & \text{if } s<s^*, \end{cases}$$

where $s_*=\beta(1-\sqrt{\gamma})^2$, $s^*=\beta(1+\sqrt{\gamma})^2$, $\sigma_\beta(\cdot)$ is the probability distribution function of the Marchenko–Pastur law specified in (4.10) and

$$\alpha_{V,\beta} = -\lim_{p\to\infty} \frac{1}{p} \ln \frac{Z^{p-1}_{pV/(p-1),\beta}}{Z^{p}_{V,\beta}}.$$

A direct calculation gives that for $p/n \to \gamma$,

$$\begin{aligned}\ln \frac{Z^{p-1}_{pV/(p-1),\beta}}{Z^{p}_{V,\beta}} &\sim \frac{\beta(n+p)}{2}\ln n - \frac{\beta p}{2}\ln p - \frac{\beta n}{2}\ln n - \frac{\beta(p+n)}{2} \times (\ln\beta - 1) + O(\ln n)\\ &\sim \frac{\beta}{2}\left\{\gamma \ln(1/\gamma) - (\gamma+1)(\ln\beta - 1)\right\} n + o(n). \qquad (4.17)\end{aligned}$$

Then, we obtain $\alpha_{V,\beta} = (\beta/2) \times \{\ln\gamma + (1/\gamma + 1)(\ln\beta - 1)\}$. Therefore, the large deviation approximation of $\lambda_1 = \max(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_p)$ has speed p and rate function

$$I_\beta(s) = \begin{cases} -\beta\int_{\mathbb{R}} \ln|s-t|\sigma_\beta(dt) + s/(2\gamma) - (\beta/2)(1/\gamma - 1)\ln s & \\ \qquad +(\beta/2)\{\ln\gamma + (1/\gamma+1)(\ln\beta - 1)\} & \text{if } s \ge s^*,\\ \infty & \text{if } s < s^*. \end{cases} \qquad (4.18)$$

Recall the notation in Remark IV.2 and from result in Bianchi et al. (2011), we know when $\mathbf{X}$ has iid entries $\mathcal{N}(0,1)$ or $\mathcal{CN}(0,1)$, the largest eigenvalue $\bar{\lambda}_1$ and the ratio $U_{n,p}$ defined in (4.1) of $\mathbf{X}^H\mathbf{X}/n$ have the same large deviation approximation function (4.18). But now in the complex case, $\mathbf{X}$ has iid entries $\mathcal{CN}(0,2)$ with $\beta = 2$. Similar to the argument in Remark IV.2, since $U_{n,p}$ is invariant to this change, we have

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\ln\Pr(U_{n,p} > x) &= \lim_{n\to\infty}\frac{1}{n}\ln\Pr(\bar{\lambda}_1 > x)\\ &= \lim_{p\to\infty}\frac{p}{n}\times\frac{1}{p}\ln\Pr(\lambda_1 > \beta x) = -\gamma I_\beta(\beta x).\end{aligned}$$

Therefore we have the large deviation result:

$$n^{-1} \ln \alpha_{n,p}(x) \sim -\gamma I_\beta(\beta x) \,. \tag{4.19}$$

Step 2. We focus on $\ln\{\mathrm{E}_Q(L_{n,p}^2)\}$ in this step. Recall that σ_β in (4.10) denotes the equilibrium measure for the large deviations of the empirical distribution of eigenvalues $(\lambda_1, \ldots, \lambda_p)$ under P; see Lemma 2.6.2 from Anderson et al. (2010). Define t_1 as a constant such that $t_1 > n/(n-1)$ but close to $n/(n-1)$. Let $B(\epsilon)$ be the ball of probability measures defined on $[0, t_1 M]$ with radius ϵ around σ_β under the following metric ρ that generates the weak convergence of probability measures on $\mathbb{R}$. For two probability measures μ and ν on $\mathbb{R}$,

$$\rho(\mu, \nu) = \sup_{\|h\|_L \le 1} \Big| \int_{\mathbb{R}} h(x)\mu(dx) - \int_{\mathbb{R}} h(x)\nu(dx) \Big|, \tag{4.20}$$

where h is a bounded Lipschitz function defined on $\mathbb{R}$ with

$$\|h\| = \sup_{x\in\mathbb{R}} |h(x)|, \quad \|h\|_L = \|h\| + \sup_{x\neq y} |h(x) - h(y)|/|x-y|.$$

Let $\mathcal{L}_{p-1}^Q$ be the empirical measure of $(\lambda_2^*, \ldots, \lambda_p^*)$ with $(\lambda_2, \ldots, \lambda_p) = \{(n-1)/n\} \times (\lambda_2^*, \ldots, \lambda_p^*)$ being constructed as in Step 1 of Algorithm IV.1 under the change of measure Q.

We know from the Marchenko–Pastur law that $\mathcal{L}_{p-1}^Q \to \sigma_\beta$ a.s., as defined in

(4.10). Then for a large constant M, we have the following upper bound for $\mathrm{E}_Q(L_{n,p}^2)$

$$
\begin{aligned}
\mathrm{E}_Q(L_{n,p}^2) \le\ & \mathrm{E}_Q\left\{(dP/dQ)^2 : \lambda_1 > M\right\} \\
& + \mathrm{E}_Q\left\{(dP/dQ)^2 : U_{n,p} > x, M > \lambda_1, \mathcal{L}_{p-1}^Q \notin B(\epsilon)\right\} \\
& + \mathrm{E}_Q\left\{(dP/dQ)^2 : U_{n,p} > x, M > \lambda_1, \mathcal{L}_{p-1}^Q \in B(\epsilon)\right\} \\
\equiv\ & I_1 + I_2 + I_3. \qquad (4.21)
\end{aligned}
$$

We will show that the first two terms of the above upper bound are ignorable, i.e., for any $\epsilon > 0$,

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln I_1 = -\infty, \qquad (4.22)$$

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln I_2 = -\infty, \qquad (4.23)$$

and we will further show that

$$\lim_{\epsilon\to 0, M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln I_3 = -2\gamma I_\beta(\beta x). \qquad (4.24)$$

Combining (4.22), (4.23) and (4.24) together, we will then deduce that

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathrm{E}_Q(L_{n,p}^2) \le -2\gamma I_\beta(\beta x).$$

Then by the result in Step 1 of the proof, and the fact that $\ln \alpha_{n,p}(x) < 0$, we will conclude that

$$\liminf_{n\to\infty} \frac{\ln \mathrm{E}_Q(L_{n,p}^2)}{2\ln \alpha_{n,p}(x)} \ge 1.$$

Based on the argument above, in the following we need only prove (4.22)–(4.24).

Proof of (4.22). Let $B_{n,p,\beta} = Z_{pV/(p-1),\beta}^{p-1}/Z_{V,\beta}^{p}$. From the construction of the change of measure Q, we can rewrite the left-hand side display in (4.22) as

$$\lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n}\ln \mathrm{E}_Q\left[\left\{\frac{B_{n,p,\beta}\prod_{i=2}^{p}(\lambda_1-\lambda_i)^{\beta}\times\lambda_1^{\beta(n-p+1)/2-1}\times e^{-n\lambda_1/2}}{nre^{-nr(\lambda_1-\tilde{x}\vee\lambda_2)}\times\mathbf{1}_{(\lambda_1>\tilde{x}\vee\lambda_2)}}\right\}^2 : \lambda_1>M\right]$$

$$\leq \lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n}\ln\int\limits_{\substack{\lambda_1>M,\\ \lambda_1>\lambda_2}} r^{-2}n^{-2}B_{n,p,\beta}^2\lambda_1^{\beta(p+n-1)-2}e^{-n\lambda_1+2rn(\lambda_1-\tilde{x}\vee\lambda_2)}$$

$$\times\, rne^{-rn(\lambda_1-\tilde{x}\vee\lambda_2)}f_{n,p}^{Q}(\lambda_2,\ldots,\lambda_p)d\lambda_1\cdots d\lambda_p$$

$$\leq \lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n}\ln\int\limits_{\lambda_1>M} r^{-1}n^{-1}B_{n,p,\beta}^2\lambda_1^{\beta(p+n-1)-2}\times e^{-n\lambda_1+rn\lambda_1-rn\tilde{x}}d\lambda_1.$$

Next we change variable λ_1 to λ_1+M, and since $(\lambda_1+M)^{\beta(p+n-1)-2}\leq M^{\beta(p+n-1)-2}\times e^{\{\beta(p+n-1)-2\}\lambda_1/M}$, we obtain the following upper bound for the expectation in Eq. (4.22):

$$\lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n}\ln\int_0^{\infty} r^{-1}n^{-1}B_{n,p,\beta}^2M^{\beta(p+n-1)-2}e^{\{\beta(p+n-1)-2\}\lambda_1/M-(n-rn)(\lambda_1+M)-rn\tilde{x}}d\lambda_1$$

$$=\lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n}\ln\{B_{n,p,\beta}^2M^{\beta(p+n-1)-2}e^{-(n-rn)M-rn\tilde{x}}\}+o(1)=-\infty,$$

where the last step follows from the approximation of $B_{n,p,\beta}$ from (4.17). This proves Eq. (4.22).

Proof of (4.23). Consider the expectation term in Eq. (4.23). Since $\lambda_1-\lambda_i<M$ and $\lambda_2\vee\tilde{x}\geq\tilde{x}$, the following inequality holds for any $\epsilon>0$,

$$\limsup_{n\to\infty}\frac{1}{n}\ln I_2\leq$$
$$\limsup_{n\to\infty}\frac{1}{n}\ln \mathrm{E}_Q\left[\left\{\frac{B_{n,p,\beta}M^{\beta(p-1)}\lambda_1^{\beta(n-p+1)/2-1}e^{-n\lambda_1/2}}{rne^{-rn(\lambda_1-\tilde{x})}}\right\}^2 : U_{n,p}>x, M>\lambda_1, \mathcal{L}_{p-1}^{Q}\notin B(\epsilon)\right]. \tag{4.25}$$

Under the assumption that $p/n\to\gamma$, $\lambda_1<M$ and with the result from (4.17), we know that

$$\frac{B_{n,p,\beta}M^{\beta(p-1)}\lambda_1^{\beta(n-p+1)/2-1}e^{-n\lambda_1/2}}{rne^{-rn(\lambda_1-\tilde{x})}}=e^{O(nM)}.$$

This implies that

$$
\begin{aligned}
(4.25) \leq\ & \limsup_{n\to\infty} n^{-1} \ln \left[e^{O(nM)} Q\{U_{n,p} > x, M > \lambda_1, \mathcal{L}^Q_{p-1} \notin B(\epsilon)\} \right] \\
\leq\ & \limsup_{n\to\infty} \left[O(M) + n^{-1} \ln \Pr\{\mathcal{L}^Q_{p-1} \notin B(\epsilon)\} \right].
\end{aligned}
$$

The large deviation result for $\mathcal{L}^Q_{p-1}$ (Theorem 2.6.1 in Anderson et al., 2010) then yields

$$
\limsup_{n\to\infty} \frac{1}{n^2} \ln \Pr\{\mathcal{L}^Q_{p-1} \notin B(\epsilon)\} = \limsup_{n\to\infty} \frac{(p-1)^2}{n^2} \times \frac{1}{(p-1)^2} \ln \Pr\{\mathcal{L}^Q_{p-1} \notin B(\epsilon)\} < 0.
$$

This proves (4.23).

Proof of (4.24) Define $\Omega_n = \{U_{n,p} > x, M > \lambda_1$ and $\mathcal{L}^Q_{p-1} \in B(\epsilon)\}$. We can write

$$
I_3 = O(1) n^{-2} B^2_{n,p,\beta} \mathrm{E}_Q \left\{ e^{2\beta \sum_{i=2}^p \ln(\lambda_1 - \lambda_i)} \lambda_1^{\beta(n-p+1)-2} e^{-n\lambda_1} e^{2nr(\lambda_1 - \tilde{x} \vee \lambda_2)} : \Omega_n \right\}.
$$

Let $\Phi(z, \epsilon) = \sup_{\mu \in B(\epsilon)} \int \ln(|z-y|)\{\mu(dy) - \sigma_\beta(dy)\}$, we have

$$
\begin{aligned}
\sum_{i=2}^p \ln(\lambda_1 - \lambda_i) &= (p-1) \int_{\mathbb{R}} \ln\left(\frac{n\lambda_1}{n-1} - y\right) \mathcal{L}^Q_{p-1}(dy) - (p-1) \ln \frac{n}{n-1} \\
&\leq (p-1)\Phi\left(\frac{n\lambda_1}{n-1}, \epsilon\right) + (p-1) \int \ln\left(\frac{n\lambda_1}{n-1} - y\right) \sigma_\beta(dy) + O(1).
\end{aligned}
$$

Under the condition that $\lambda_1 < M$, we know $n\lambda_1/(n-1) < 2M$ when n is large enough. Let $G = \max\{\beta(1+\sqrt{\gamma})^2, 2M\}$ and define

$$
h(x) = x\mathbf{1}(x \in [0, G]). \tag{4.26}
$$

Then h is a bounded Lipschitz function on $[0, G]$. Furthermore, given $\mathcal{L}^Q_{p-1} \in B(\epsilon)$

and under measure Q, we have

$$\Big|\frac{1}{p-1}\sum_{i=2}^{p}\frac{n\lambda_i}{n-1}-\beta\Big|=\Big|\int\limits_{\mathbb{R}} h(y)\mathcal{L}_{p-1}^{Q}(dy)-\int\limits_{\mathbb{R}} h(y)\sigma_\beta(dy)\Big|<O(\epsilon),$$

for $\beta\in\{1,2\}$. This is because from Theorem 6.3.1 in Dumitriu and Edelman (2003), for a distribution with the same density as (4.13), the first moment is $\mu_{1,\gamma}=\int \bar{s}\times f(\bar{s})d\bar{s}=1$. For the density in (4.10), similar to Remark IV.2, the first moment is $\int s\times\sigma_\beta(ds)=\beta\int\bar{s}\times f(\bar{s})d\bar{s}=\beta\times\mu_{1,\gamma}=\beta$. Considering our choice of G in (4.26), we have $\int_{\mathbb{R}} h(y)\sigma_\beta(dy)=\int_{\mathbb{R}} y\sigma_\beta(dy)=\beta\times\mu_{1,\gamma}=\beta$. Therefore, $U_{n,p}>x$ and $\lambda_1>\tilde{x}$ implies that $\lambda_1>\beta x+O(\epsilon)$ and we can write

$$I_3\le O(1)n^{-1}B_{n,p,\beta}^2\int\limits_{\beta x+O(\epsilon)}^{M} e^{2\beta(p-1)\Phi(\frac{n\lambda_1}{n-1},\epsilon)+2\beta(p-1)\int\ln(\frac{n\lambda_1}{n-1}-y)\sigma_\beta(dy)}$$
$$\times\lambda_1^{\beta(n-p+1)-2}e^{-n\lambda_1+rn\{\lambda_1-\beta x+O(\epsilon)\}}d\lambda_1.$$

Since $\beta x+O(\epsilon)<\lambda_1<M$, we have

$$\Phi[n\lambda_1/(n-1),\epsilon]\le\sup_{z\in[n\{\beta x+O(\epsilon)\}/(n-1),nM/(n-1)]}\Phi(z,\epsilon)$$

under the constraint $\mathcal{L}_{p-1}^{Q}\in B(\epsilon)$ and that

$$\begin{aligned}&\int\ln\left(\frac{n\lambda_1}{n-1}-y\right)\sigma_\beta(dy)\\ =&\int\ln\left(\frac{n\beta x}{n-1}-y\right)\sigma_\beta(dy)+\int\ln\left\{1+\frac{n\lambda_1-n\beta x}{n\beta x-(n-1)y}\right\}\sigma_\beta(dy)\\ \le&\int\ln\left(\frac{n\beta x}{n-1}-y\right)\sigma_\beta(dy)+\int\frac{n\lambda_1-n\beta x}{n\beta x-(n-1)y}\sigma_\beta(dy).\end{aligned}$$

It follows that

$$I_3 \le O(1) n^{-1} B^2_{n,p,\beta} \times e^{2\beta(p-1)\sup_{z\in[\frac{n(\beta x+O(\epsilon))}{n-1},\frac{nM}{n-1}]}\Phi(z,\epsilon)+2\beta(p-1)\int \ln(\frac{n\beta x}{n-1}-y)\sigma_\beta(dy)}$$
$$\times \int_{\beta x+O(\epsilon)}^{M} e^{2\beta(p-1)\int \frac{n\lambda_1-n\beta x}{n\beta x-(n-1)y}d\sigma_\beta(y)} \lambda_1^{\beta(n-p+1)-2} e^{-n\lambda_1+rn\{\lambda_1-\beta x+O(\epsilon)\}} d\lambda_1.$$

The right-hand side equals

$$O(1) n^{-1} B^2_{n,p,\beta} \times e^{2\beta(p-1)\sup_{z\in[\frac{n(\beta x+O(\epsilon))}{n-1},\frac{nM}{n-1}]}\Phi(z,\epsilon)+2\beta(p-1)\int \ln(\frac{n\beta x}{n-1}-y)\sigma_\beta(dy)}$$
$$\times \int_{O(\epsilon)}^{M-\beta x} e^{2\beta(p-1)\int \frac{n\lambda_1}{n\beta x-(n-1)y}d\sigma_\beta(y)} \times (\lambda_1+\beta x)^{\beta(n-p+1)-2} \times e^{-(1-r)n(\lambda_1+\beta x)-rn\{\beta x+O(\epsilon)\}} d\lambda_1,$$

where we change the variable λ_1 to $\lambda_1 + \beta x$ for the integral. Then it follows that

$$I_3 \le O(1) n^{-1} B^2_{n,p,\beta} \times e^{2\beta(p-1)\sup_{z\in[\frac{n(\beta x+O(\epsilon))}{n-1},\frac{nM}{n-1}]}\Phi(z,\epsilon)+2\beta(p-1)\int \ln(\frac{n\beta x}{n-1}-y)\sigma_\beta(dy)}$$
$$\times \ (\beta x)^{\beta(n-p+1)-2} e^{-n\{\beta x+O(\epsilon)\}}$$
$$\times \int_{O(\epsilon)}^{M-\beta x} e^{2\beta(p-1)\int \frac{n\lambda_1}{n\beta x-(n-1)y}d\sigma_\beta(y)+\{\beta(n-p+1)-2\}\frac{\lambda_1}{\beta x}-(1-r)n\lambda_1} d\lambda_1, \tag{4.27}$$

as we used the fact that $(\lambda_1 + \beta x)^{\beta(n-p+1)-2} \le (\beta x)^{\beta(n-p+1)-2} e^{\{\beta(n-p+1)-2\}\lambda_1/(\beta x)}$.

Under $s^* < \beta x$, we can find a finite number t_0 such that $s^* < t_0 x \le \{\beta x + O(\epsilon)\} \times n/(n-1)$, for small enough ϵ and large enough n. Recall that $t_1 M \ge nM/(n-1)$. Next we show that

$$\limsup_{\epsilon\to 0} \sup_{z\in[t_0 x, t_1 M]} \Phi(z,\epsilon) \le 0. \tag{4.28}$$

For any $z \in [t_0 x, t_1 M]$ and $\mu \in B(\epsilon)$, let $\mathcal{S}_1(z) = \{y \in \mathrm{supp}(\sigma_\beta) \cup \mathrm{supp}(\mu) : |z-y| > \eta\}$ and $\mathcal{S}_2(z) = \{y \in \mathrm{supp}(\sigma_\beta) \cup \mathrm{supp}(\mu) : |z-y| \le \eta\}$, where $\mathrm{supp}(\mu)$ is the support of measure μ and η is a small constant such that $\eta < \min\{t_0 x - s^*, 1\}$ with s^* defined in (4.10). Note that $\mathrm{supp}(\sigma_\beta) \subset \mathcal{S}_1(z)$. Given $z \in [t_0 x, t_1 M]$, set $f_z(y) = \ln(|z-y|)$ for $y \in \mathcal{S}_1(z)$. The Lipschitz norms of the set of functions $\{f_z(\cdot);\, z \in [t_0 x, t_1 M]\}$ on $\mathcal{S}_1(z)$ are bounded by a constant $C < \infty$. By the definition of $\rho(\cdot,\cdot)$ in (4.20), we obtain

$$
\begin{aligned}
&\sup_{z\in[t_0x,t_1M]} \int_{\mathbb{R}} \ln(|z-y|)\{\mu(dy) - \sigma_\beta(dy)\} \\
\le &\sup_{z\in[t_0x,t_1M]} \int_{\mathcal{S}_1} f_z(y)\{\mu(dy) - \sigma_\beta(dy)\} + \sup_{z\in[t_0x,t_1M]} \int_{\mathcal{S}_2} f_z(y)\mu(dy) \\
\le &\sup_{z\in[t_0x,t_1M]} \int_{\mathcal{S}_1} f_z(y)\{\mu(dy) - \sigma_\beta(dy)\} \le C\rho(\mu,\sigma_\beta) < C\epsilon,
\end{aligned}
$$

for any $\mu \in B_\epsilon$. This implies that $\sup_{z\in[t_0x,t_1M]} \Phi(z,\epsilon) < C\epsilon$. Then (4.28) follows. When $r < 1 - 2\beta\gamma \int \{1/(\beta x - y)\} d\sigma_\beta(y) - (1-\gamma)/x$, we know that the integral term in (4.27) is $\sim e^{nO(\epsilon)}$. Therefore

$$
\begin{aligned}
&\lim_{\substack{\epsilon\to 0\\ M\to\infty}} \limsup_{n\to\infty} \frac{1}{n} \ln I_3 \\
=&2\beta\gamma \int \ln(\beta x - y)\sigma_\beta(dy) - \beta x + \beta(1-\gamma)\ln(\beta x) - \beta\left\{\gamma \ln\gamma + (1+\gamma)(\ln\beta - 1)\right\} \\
=&-2\gamma I_\beta(\beta x),
\end{aligned}
$$

where $I_\beta(x)$ is defined as in (4.18). Therefore, we obtain $\limsup_{n\to\infty} \ln \mathrm{E}_Q(L_{n,p}^2)/n \le -2\gamma I_\beta(\beta x)$. Hence, the above upper bound and the approximation in (4.19) imply that $\liminf_{n\to\infty} \ln \mathrm{E}_Q(L_{n,p}^2)/\{2\ln\alpha_{n,p}(x)\} \ge 1$, where note that $\ln\alpha_{n,p}(x) < 0$. This completes the proof.

APPENDICES

APPENDIX A

Appendix of Chapter II

This Appendix provides the proofs of theoretical results in Chapter II. In particular, Section A.1 present the proofs of theoretical results in Section 2.2, Section A.2 proves the technical lemmas in Section A.1, Section A.3 provides the proofs of theoretical results in Section 2.3, and Section A.4 the proofs of theoretical results in Section 2.4.

A.1 Proofs of Theoretical Results in Section 2.2

Section A.1.1 presents the proofs for the testing problem (III) as an illustration example, where the problem jointly testing the the one-sample mean vector and covariance matrix. Other testing problems (I)–(II) and (IV)–(VII) can be proved following a similar analysis, and are discussed in Sections A.1.2–A.1.5. In this section, we let $a_n \sim b_n$ denote $\lim_{n\to\infty} |a_n/b_n| = 1$.

A.1.1 Proof Illustration: Theorems 2.2.1–2.2.3 (III)

A.1.1.1 Proof of Theorem 2.2.1 (III)

When p is fixed, the chi-squared approximations hold by the classical multivariate analysis (Anderson, 2003; Muirhead, 2009). Therefore, without loss of generality, the proofs below focus on $p \to \infty$.

Deriving the necessary and sufficient conditions for the chi-squared approximations requires the correct understanding of the limiting behavior of $\log \Lambda_n$ under both low and high dimensions. Particularly, we examine the limiting distribution of the log likelihood ratio test statistic $\log \Lambda_n$ based on the moment generating function of $\log \Lambda_n$, that is, $\mathrm{E}\{\exp(t \log \Lambda_n)\}$. For Λ_n in question (III), by Theorem 8.5.3 and Corollary 8.5.4 in Muirhead (2009), we have that under H_0,

$$\mathrm{E}\{\exp(t \log \Lambda_n)\} = \mathrm{E}(\Lambda_n^t) = \left(\frac{2e}{n}\right)^{npt/2} (1+t)^{-np(1+t)/2} \times \frac{\Gamma_p[\{n(1+t)-1\}/2]}{\Gamma_p\{(n-1)/2\}}, \tag{A.1}$$

where $\Gamma_p(\cdot)$ is the multivariate Gamma function; see Definition 2.1.10 in Muirhead (2009).

When p is fixed, the moment generating function of $-2\log \Lambda_n$ approximates that of a chi-squared variable χ_f^2, where $f = p(p+3)/2$; see, Sections 8.2.4 and 8.5 in Muirhead (2009). When $p \to \infty$, Jiang and Yang (2013) and Jiang and Qi (2015) derived an approximate expansion of the multivariate Gamma function, and their Theorem 5 utilized (A.1) to show that under the conditions of Theorem 2.2.1,

$$\mathrm{E}[\exp\{s(-2\log \Lambda_n + 2\mu_n)/(2n\sigma_n)\}] \to \exp(s^2/2), \tag{A.2}$$

where $\exp(s^2/2)$ is the moment generating function of $\mathcal{N}(0,1)$, and

$$\mu_n = -\frac{1}{4}\left\{n(2n-2p-3)\log\left(1-\frac{p}{n-1}\right) + 2(n+1)p\right\}, \tag{A.3}$$

$$\sigma_n^2 = -\frac{1}{2}\left\{\frac{p}{n-1} + \log\left(1-\frac{p}{n-1}\right)\right\}. \tag{A.4}$$

We next prove (i) in Theorem 2.2.1 when $p \to \infty$ based on (A.2). Particularly, we write

$$\begin{aligned}
& \sup_{\alpha\in(0,1)} \left|\Pr\{-2\log\Lambda_n > \chi_f^2(\alpha)\} - \alpha\right| \\
= & \sup_{\alpha\in(0,1)} \left|\Pr(T_n > q_{n,\alpha}) - \bar\Phi(q_{n,\alpha}) + \bar\Phi(q_{n,\alpha}) - \bar\Phi(z_\alpha)\right|, \qquad \text{(A.5)}
\end{aligned}$$

where $T_n = (-2\log\Lambda_n + 2\mu_n)/(2n\sigma_n)$, $q_{n,\alpha} = \{\chi_f^2(\alpha) + 2\mu_n\}/(2n\sigma_n)$, and $\bar\Phi(\cdot) = 1 - \Phi(\cdot)$ with $\Phi(\cdot)$ being the cumulative distribution function of $\mathcal{N}(0,1)$. Since (A.2) suggests that T_n converges to $\mathcal{N}(0,1)$ in distribution, and the cumulative distribution function of $\mathcal{N}(0,1)$ is continuous, by Pólya-Cantelli Lemma (see, e.g., Lemma 2.11 in Van der Vaart (2000)), we have $\sup_{\alpha\in(0,1)} |\Pr(T_n > q_{n,\alpha}) - \bar\Phi(q_{n,\alpha})| \to 0$. Consequently, (A.5) $\to 0$ if and only if $\sup_{\alpha\in(0,1)} |\bar\Phi(q_{n,\alpha}) - \bar\Phi(z_\alpha)| \to 0$, which is equivalent to $\sup_{\alpha\in(0,1)} |q_{n,\alpha} - z_\alpha| \to 0$, as $\bar\Phi(\cdot)$ is a continuous and strictly decreasing function with bounded derivative. Since χ_f^2 can be viewed as a summation over f independent χ_1^2 variables, and $f \to \infty$ as $p \to \infty$, we can apply Berry–Esseen theorem to χ_f^2 variable, and obtain

$$\sup_{\alpha\in(0,1)} \left|\{\chi_f^2(\alpha) - f\}/\sqrt{2f} - z_\alpha\right| = O(f^{-1/2}). \qquad \text{(A.6)}$$

Therefore, $\sup_{\alpha\in(0,1)} |q_{n,\alpha} - z_\alpha| \to 0$ is equivalent to

$$\sqrt{2f} \times (2n\sigma_n)^{-1} \to 1, \qquad \text{(A.7)}$$

$$(O(1) + f + 2\mu_n) \times (2n\sigma_n)^{-1} \to 0. \qquad \text{(A.8)}$$

Following similar analysis, we know that under the conditions of Theorem 2.2.1, and when $p \to \infty$, for the chi-squared approximation with the Bartlett correction,

$\sup_{\alpha\in(0,1)} |\Pr\{-2\rho\log\Lambda_n > \chi_f^2(\alpha)\} - \alpha|$ holds if and only if

$$\sqrt{2f} \times (2n\rho\sigma_n)^{-1} \to 1, \tag{A.9}$$

$$(O(1) + f + 2\rho\mu_n) \times (2n\rho\sigma_n)^{-1} \to 0. \tag{A.10}$$

We next examine (A.7)–(A.8) and (A.9)–(A.10) for the chi-squared approximation without and with the Bartlett correction, respectively.

(III.i) The chi-squared approximation. We next discuss two cases $\lim_{n\to\infty} p/n = 0$ and $\lim_{n\to\infty} p/n = C \in (0,1]$, respectively.

Case (III.i.1) $\lim_{n\to\infty} p/n = 0$. Under this case, we prove that (A.7) holds. As $\sqrt{2f} \sim p$, it is equivalent to show that $p/(2n\sigma_n) \to 1$. By Taylor's expansion of σ_n^2 in (A.4), we have

$$2\sigma_n^2 = -\frac{p}{n-1} - \log\left(1 - \frac{p}{n-1}\right) = \frac{p^2}{2(n-1)^2} + o\left(\frac{p^2}{n^2}\right),$$

and therefore $\sqrt{2f} \times (2n\sigma_n)^{-1} \to 1$. We next show that (A.8) holds if and only if $p^2/n \to 0$. Given (A.7) and $\sqrt{2f} \sim p$, (A.8) is equivalent to $(f + 2\mu_n)/p \to 0$. By $p/n = o(1)$ and Taylor's expansion of $\log(1-x)$, for μ_n in (A.3), we have

$$\begin{aligned} 4\mu_n/p =& -2(n+1) + n(2n-2p-3)\left\{\frac{1}{n-1} + \frac{p}{2(n-1)^2} + \frac{p^2}{3(n-1)^3} + O\left(\frac{p^3}{n^4}\right)\right\} \\ =& -2p - 3 + \frac{(2n-2p-3)p}{2(n-1)} + \frac{(2n-2p-3)p^2}{3(n-1)^2} + o(1) + O\left(\frac{p^3}{n^2}\right). \end{aligned} \tag{A.11}$$

As $2f/p = p + 3$, we obtain

$$\begin{aligned} 2(f+2\mu_n)/p =& -p + \frac{\{2(n-1) - 2p - 1\}p}{2(n-1)} + \frac{2p^2}{3(n-1)} + o(1) + O\left(\frac{p^3}{n^2}\right) \\ =& -\frac{p^2}{3(n-1)} + o(1) + O\left(\frac{p^3}{n^2}\right). \end{aligned} \tag{A.12}$$

Therefore when $p/n \to 0$, (A.8) holds if and only if $p^2/n \to 0$.

Case (III.i.2) $\lim_{n\to\infty} p/n = C \in (0,1]$. Under this case, we have

$$\sqrt{2f} \times (2n\sigma_n)^{-1} \sim p(2n\sigma_n)^{-1} \sim C(2\sigma_n)^{-1}. \tag{A.13}$$

If $C = 1$, $\sigma_n^2 \to \infty$ and thus (A.13) $\to 0$. If $C \in (0,1)$, we have $C(2\sigma_n)^{-1} \sim C[-2\{C + \log(1-C)\}]^{-1/2} < 1$ when $0 < C < 1$. In summary, (A.7) does not hold, which suggests that the chi-squared approximation fails.

Finally, we consider a general sequence $p/n = p_n/n \in [0,1]$, where we write p as p_n to emphasize that p changes with n. Similarly, we also write f as f_n. Note that a sequence converges if and only if every subsequence converges. For the sequence $\{p_n/n\}$, by the Bolzano–Weierstrass theorem, we can further take a subsequence $\{n_t\}$ such that $p_{n_t}/n_t \to C \in [0,1]$. If $C \in (0,1]$, the above analysis still applies, which shows that the chi-squared approximation fails. Alternatively, if all the subsequences of $\{p/n\}$ converge to 0, we know $p/n \to 0$. In summary, the above analysis shows that (A.7) and (A.8) hold if and only if $p^2/n \to 0$.

(III.ii) The chi-squared approximation with the Bartlett correction. Similarly to the analysis above, we discuss two cases $\lim_{n\to\infty} p/n = 0$ and $\lim_{n\to\infty} p/n = C \in (0,1]$, respectively.

Case (III.ii.1) $\lim_{n\to\infty} p/n = 0$. Under this case, we know (A.9) holds since $\rho = 1 + O(p/n) \to 1$ and $p/(2n\sigma_n) \to 1$ as shown in Case (III.i.1) above. Given (A.9), deriving the condition for (A.10) is equivalent to examine when $p^{-1}(f + 2\rho\mu_n) \to 0$. Following the analysis of (A.12), we further obtain

$$\begin{aligned} 2(f+2\mu_n)/p =& (p+3) - 2(n+1) + n(2n-2p-3)\sum_{j=1}^{4} \frac{p^{j-1}}{j(n-1)^j} + O\left(\frac{p^4}{n^3}\right) \\ =& -\frac{p^2}{3(n-1)} - \frac{p^3}{6(n-1)^2} + O\left(\frac{p^4}{n^3}\right) + o(1). \end{aligned} \tag{A.14}$$

We write $\rho = 1 - \Delta_n$ where $\Delta_n = \{6n(p+3)\}^{-1}(2p^2+9p+11)$, which is $O(p/n)$. By (A.12), we have $4\mu_n/p = -p - 3 - p^2/\{3(n-1)\} + o(1) + O(p^3 n^{-2})$. Together with (A.14), we have

$$\begin{aligned} & 2\times(f+2\rho\mu_n)/p = 2\times(f+2\mu_n)/p - 4\Delta_n\times\mu_n/p \qquad\qquad (A.15)\\ &= -\frac{p^2}{3(n-1)} - \frac{p^3}{6(n-1)^2} - \Delta_n\left\{-p-3-\frac{p^2}{3(n-1)}\right\} + O\left(\frac{p^4}{n^3}\right) + o(1)\\ &= -\frac{p^2}{3(n-1)} - \frac{p^3}{6(n-1)^2} + \frac{2p^2(p+3)}{6n(p+3)} + \frac{2p^2\times p^2}{6n(p+3)\times 3(n-1)} + O\left(\frac{p^4}{n^3}\right) + o(1)\\ &= -\frac{p^3}{18n^2} + O\left(\frac{p^4}{n^3}\right) + o(1). \end{aligned}$$

Therefore under this case (A.10) holds if and only if $p^3/n^2 \to 0$.

Case (III.ii.2): When $\lim_{n\to\infty} p/n = C \in (0,1]$, we have $\rho \to 1 - C/3$ and

$$\sqrt{2f}\times(2n\rho\sigma_n)^{-1} \sim C\times(1-C/3)^{-1}(2\sigma_n)^{-1}.$$

Similarly to the Case (III.i.2) above, if $C = 1$, (A.9) $\to 0$; if $C \in (0,1)$, we have $C(1-C/3)^{-1}(2\sigma_n)^{-1} \sim C(1-C/3)^{-1}[-2\{C+\log(1-C)\}]^{-1/2} < 1$ when $0 < C < 1$. In summary, (A.9) does not hold, which suggests the failure of the chi-squared approximation with the Bartlett correction.

For a general sequence $p/n = p_n/n \in [0,1]$, the analysis of taking subsequences above can be applied similarly. In summary, we know that for the likelihood ratio test in problem (III), the chi-squared approximation with the Bartlett correction holds if and only if $p^3/n^2 \to 0$.

A.1.1.2 Proof of Theorem 2.2.2 (III)

We prove Theorem 2.2.2 for problem (III) by examining the characteristic function of $-2\eta\log\Lambda_n$, where $\eta = 1$ or $\eta = \rho$, and ρ is the corresponding Bartlett correction factor, given in Section 2.2.1. The following Lemma A.1.1 gives an asymptotic expansion for the characteristic function $\mathrm{E}\{\exp(-2it\eta\log\Lambda_n)\}$, where the notation i is reserved to denote the solution of the equation $x^2=-1$, i.e., the imaginary unit.

Lemma A.1.1. *Under H_0 of the testing problem (III), when $\eta=1$ or $\eta=\rho$ with the Bartlett correction factor ρ in Section 2.2.1, the characteristic function of $-2\eta\log\Lambda_n$ satisfies that for a given integer L, when $p^{L+2}/n^L\to 0$,*

$$\mathrm{E}\{\exp(-2it\eta\log\Lambda_n)\}=(1-2it)^{-f/2}\exp\left[\sum_{l=1}^{L-1}\varsigma_l\big\{(1-2it)^{-l}-1\big\}+O\bigg(\frac{p^{L+2}}{n^L}\bigg)\right],$$

where $f=p(p+3)/2$ is the corresponding degrees of freedom, and

$$\varsigma_l=\frac{(-1)^{l+1}}{l(l+1)}\sum_{j=1}^{p}\left\{B_{l+1}\left(\frac{(1-\eta)n}{2}-\frac{j}{2}\right)-\left(\frac{(1-\eta)n}{2}\right)^{l+1}\right\}\left(\frac{\eta n}{2}\right)^{-l}. \qquad \text{(A.16)}$$

For any integer $l\geq 1$, $B_l(\cdot)$ represents the Bernoulli polynomial of degree l; see, e.g., Eq. (25) in Section 8.2.4 of Muirhead (2009).

Proof. Section A.2.2.1 on Page 188. $\square$

With Lemma A.1.1, we next prove (2.1) and (2.2) in Theorem 2.2.2 for the chi-squared approximations without and with the Bartlett correction, respectively.

(i) The chi-squared approximation. When $\rho=1$, as $B_{l+1}(\cdot)$ is a polynomial of order $l+1$, we have $\varsigma_l=O(p^{l+2}n^{-l})$ for $l\geq 2$, and we can check that $\varsigma_1=\Theta(p^3n^{-1})$; see (A.23). Thus when $p^2/n\to 0$, $\varsigma_l\to 0$ for $l\geq 2$. Let $\Psi(t)=\mathrm{E}\{\exp(-2it\log\Lambda_n)\}$. Then by Lemma A.1.1,

$$\Psi(t)=(1-2it)^{-f/2}\left\{\exp\left[\sum_{l=1}^{2}\varsigma_l\big\{(1-2it)^{-l}-1\big\}+O\big(p^5n^{-3}\big)\right]\right\}. \qquad \text{(A.17)}$$

By Taylor' expansion, we can write $\exp[\varsigma_l\{(1-2it)^{-l}-1\}] = 1 + V_l(t)$, where

$$V_l(t) = \sum_{v=1}^{\infty} \frac{\varsigma_l^v}{v!} \sum_{w=0}^{v} \binom{v}{w} (1-2it)^{-lw}(-1)^{v-w}. \tag{A.18}$$

Then by (A.17) and $p^2/n \to 0$, we have $\Psi(t) = \tilde{\Psi}(t)\{1 + O(p^5/n^3)\}$, where

$$\begin{aligned}
\tilde{\Psi}(t) &= (1-2it)^{-f/2}\big\{1 + V_1(t)\big\}\big\{1 + V_2(t)\big\} \\
&= (1-2it)^{-f/2} + \sum_{v=1}^{\infty} \frac{\varsigma_1^v}{v!} \sum_{w=0}^{v} \binom{v}{w} (1-2it)^{-f/2-w}(-1)^{v-w} \qquad \text{(A.19)} \\
&\quad + \sum_{v=1}^{\infty} \frac{\varsigma_2^v}{v!} \sum_{w=0}^{v} \binom{v}{w} (1-2it)^{-f/2-2w}(-1)^{v-w} \\
&\quad + \sum_{\substack{v_1 \geq 1;\ 0 \leq w_1 \leq v_1 \\ v_2 \geq 1;\ 0 \leq w_2 \leq v_2}} \frac{\varsigma_1^{v_1}\varsigma_2^{v_2}}{v_1! v_2!} \binom{v_1}{w_1}\binom{v_2}{w_2} (1-2it)^{-f-w_1-2w_2}(-1)^{v_1-w_1+v_2-w_2}.
\end{aligned}$$

Note that $(1-2it)^{-f/2}$ is the characteristic function of χ_f^2 distribution. Following similar analysis to Section 8.5 in Anderson (2003), we use the inversion property of the characteristic function, and then by (A.19), we obtain that

$$\begin{aligned}
&\Pr(-2\log\Lambda_n \leq x) \qquad \text{(A.20)} \\
&= \Bigg\{ \Pr(\chi_f^2 \leq x) + \sum_{v=1}^{\infty} \frac{\varsigma_1^v}{v!} \sum_{w=0}^{v} \binom{v}{w} \Pr(\chi_{f+2w}^2 \leq x)(-1)^{v-w} \\
&\quad + \sum_{v=1}^{\infty} \frac{\varsigma_2^v}{v!} \sum_{w=0}^{v} \binom{v}{w} \Pr(\chi_{f+4w}^2 \leq x)(-1)^{v-w} \\
&\quad + \sum_{\substack{v_1 \geq 1;\ 0 \leq w_1 \leq v_1 \\ v_2 \geq 1;\ 0 \leq w_2 \leq v_2}} \frac{\varsigma_1^{v_1}\varsigma_2^{v_2}}{v_1! v_2!} \binom{v_1}{w_1}\binom{v_2}{w_2} \Pr(\chi_{2f+2w_1+4w_2}^2 \leq x)(-1)^{v_1-w_1+v_2-w_2} \Bigg\} \\
&\quad \times \left\{1 + O\left(\frac{p^5}{n^3}\right)\right\}.
\end{aligned}$$

(From (A.19) to (A.20), Fubini's theorem is implicitly used to exchange the order of the infinite sum and the integration of characteristic functions.)

We next utilize the following Propositions A.1.1 and A.1.2 to evaluate (A.20).

Proposition A.1.1. *Given an integer $h \in \{1,2,3,4\}$, when $x = \chi_f^2(\alpha)$, there exists a constant C such that as $f \to \infty$,*

$$\sum_{w=0}^{v} \binom{v}{w} \Pr(\chi^2_{f+2hw} \leq x)(-1)^{v-w} = O(v! C^v f^{-v/2}) \tag{A.21}$$

uniformly over $v \geq 1$.

Proof. Please see Section A.2.2.4 on Page 193. □

Proposition A.1.2. *For $(h_1, h_2) = (1, 2)$ or $(h_1, h_2) = (2, 3)$, when $x = \chi_f^2(\alpha)$, there exists a constant C such that as $f \to \infty$,*

$$\sum_{w_1=0}^{v_1} \sum_{w_2=0}^{v_2} \binom{v_1}{w_1} \binom{v_2}{w_2} \Pr(\chi^2_{2f+2h_1w_1+2h_2w_2} \leq x)(-1)^{v_1-w_1+v_2-w_2}$$
$$= O\{v_1! v_2! C^{v_1+v_2} f^{-(v_1+v_2)/2}\}$$

uniformly over $v_1, v_2 \geq 1$.

Proof. Please see Section A.2.2.5 on Page 201. □

Remark A.1. *In Propositions A.1.1 and A.1.2, C denotes a universal constant and its value can change. This is similarly used in the following proofs. In addition, for a series $\{b_{v,f}\}$ that depends on positive integers v and f, we say $b_{v,f} = O(v! C^v f^{-v/2})$ as $f \to \infty$ and uniformly over $v \geq 1$, if there exists a constant C such that $\sup_{v \geq 1} \limsup_{f \to \infty} |b_{v,f}/(v! C^v f^{-v/2})| < \infty$.*

When $x = \chi_f^2(\alpha)$ and $f \to \infty$, we apply Proposition A.1.1 with $h = 1$ and $h = 2$, and Proposition A.1.2 with $(h_1, h_2) = (1, 2)$ to (A.20). Then as $\varsigma_1 = \Theta(p^3 n^{-1})$, $\varsigma_2 = O(p^4 n^{-2})$, and $f = \Theta(p^2)$, when $p \to \infty$ and $p^2/n \to 0$, we obtain

$$\Pr(-2 \log \Lambda_n \leq x) = \Pr(\chi_f^2 \leq x) + \varsigma_1 \left\{ \Pr(\chi^2_{f+2} \leq x) - \Pr(\chi_f^2 \leq x) \right\} + o(p^2/n). \tag{A.22}$$

We next compute ς_1. Particularly, for the chi-squared approximation, $\rho = 1$, and then by (A.16),

$$\varsigma_1 = \frac{1}{2}\sum_{j=1}^{p} B_2\left(-\frac{j}{2}\right)\left(\frac{n}{2}\right)^{-1} = \frac{1}{24n}p\left(2p^2 + 9p + 11\right), \tag{A.23}$$

where we use $B_2(z) = z^2 - z + 1/6$; see, e.g., Eq. (26) in Section 8.2.4 of Muirhead (2009). To finish the proof of (2.1), we use the following lemma.

Lemma A.1.2. *When $x = \chi_f^2(\alpha)$ and $f \to \infty$, for $h \in \{1, 2, 3, 4\}$,*

$$\Pr(\chi^2_{f+2h} \leq x) - \Pr(\chi^2_f \leq x) = -\sum_{k=1}^{h}\left\{\Gamma\left(\frac{f}{2} + h - k + 1\right)\right\}^{-1}\left(\frac{x}{2}\right)^{\frac{f}{2}+h-k} e^{-x/2} \tag{A.24}$$

$$= -\frac{h}{\sqrt{f\pi}}\exp\left(-\frac{z_\alpha^2}{2}\right)\left\{1 + O(f^{-1/2})\right\}. \tag{A.25}$$

Proof. Please see Section A.2.2.3 on Page 191. □

As $p \to \infty$, $f \to \infty$. Then by (A.22) and (A.23), and applying Lemma A.1.2 with $h = 1$, (2.1) is proved, where $\vartheta_1(n, p) = \varsigma_1/\sqrt{f}$.

(ii) The chi-squared approximation with the Bartlett correction. Similarly to the proof in Part (i) above, we prove (2.2) by examining the expansion of the characteristic function in Lemma A.1.1. In particular, for the chi-squared approximation with the Bartlett correction, we note that the Bartlett correction factor ρ is chosen such that $\varsigma_1 = 0$ (see Section 8.5.3 in Muirhead (2009)). This can be checked by plugging $\rho = 1 - \{6n(p+3)\}^{-1}(2p^2 + 9p + 11)$ into (A.16) to calculate ς_1. In addition, by $B_3(z) = z^3 - 3z^2/2 + z/2$ (see, e.g., Eq. (26) in Section 8.2.4 of Muirhead (2009)), we calculate

$$\varsigma_2 = \frac{p(2p^4 + 18p^3 + 49p^2 + 36p - 13)}{288(p+3)(\rho n)^2}, \tag{A.26}$$

and therefore $\varsigma_2 = \Theta(p^4 n^{-2})$. We redefine $\Psi(t) = \mathrm{E}\{\exp(-2it\rho \log \Lambda_n)\}$. Then when $p^3/n^2 \to 0$, by Lemma A.1.1, we have

$$\Psi(t) = (1-2it)^{-f/2}\left\{ \exp\left[\sum_{l=2}^{3} \varsigma_l \big\{(1-2it)^{-l} - 1\big\} + O(p^6 n^{-4}) \right] \right\}, \tag{A.27}$$

where we use $\varsigma_1 = 0$. Similarly to (A.19), we have $\Psi(t) = (1-2it)^{-f/2}\{1+V_2(t)\}\{1+V_3(t)\}\{1+O(p^6 n^{-4})\}$. Moreover, similarly to (A.20), we obtain

$$\begin{aligned}
&\Pr(-2\rho \log \Lambda_n \le x) \qquad\qquad (\text{A.28})\\
=&\left\{ \Pr(\chi_f^2 \le x) + \sum_{v=1}^{\infty} \frac{\varsigma_2^v}{v!} \sum_{w=0}^{v} \binom{v}{w} \Pr(\chi^2_{f+4w} \le x)(-1)^{v-w} \right.\\
&+ \sum_{v=1}^{\infty} \frac{\varsigma_3^v}{v!} \sum_{w=0}^{v} \binom{v}{w} \Pr(\chi^2_{f+6w} \le x)(-1)^{v-w}\\
&\left. + \sum_{\substack{v_2 \ge 1;\ 0 \le w_2 \le v_2 \\ v_3 \ge 1;\ 0 \le w_3 \le v_3}} \frac{\varsigma_2^{v_2} \varsigma_3^{v_3}}{v_2! v_3!} \binom{v_2}{w_2} \binom{v_3}{w_3} \Pr(\chi^2_{2f+4w_2+6w_3} \le x)(-1)^{v_2 - w_2 + v_3 - w_3} \right\}\\
&\times \left\{ 1 + O\Big(\frac{p^6}{n^4}\Big) \right\}.
\end{aligned}$$

When $x = \chi_f^2(\alpha)$ and $f \to \infty$, we apply Proposition A.1.1 with $h = 2$ and $h = 3$, and Proposition A.1.2 with $(h_1, h_2) = (2, 3)$ to (A.28). Then as $\varsigma_2 = \Theta(p^4/n^2)$, $\varsigma_3 = O(p^5/n^3)$, and $f = \Theta(p^2)$, we know that when $p \to \infty$ and $p^3/n^2 \to 0$,

$$\Pr(-2\rho \log \Lambda_n \le x) = \Pr(\chi_f^2 \le x) + \varsigma_2 \left\{ \Pr(\chi^2_{f+4} \le x) - \Pr(\chi_f^2 \le x) \right\} + o(p^3/n^2). \tag{A.29}$$

By (A.26) and (A.29), and applying Lemma A.1.2 with $h = 2$, we prove (2.2), where $\vartheta_2(n, p) = 2\varsigma_2/\sqrt{f}$.

A.1.1.3 Proof of Theorem 2.2.3 (III)

In this section, we prove Theorem 2.2.3 also by examining the characteristic function of the likelihood ratio test statistic. In particular, motivated by the limit in (A.2), we study the standardized test statistic $(-2\log\Lambda_n + 2\mu_n)(2n\sigma_n)^{-1}$, where the values of μ_n and σ_n are given in Theorem 2.2.3. Under H_0 of the testing problem (III), by (A.1), the characteristic function of $(-2\log\Lambda_n + 2\mu_n)/(2n\sigma_n)$ is

$$
\begin{aligned}
&\mathrm{E}\left\{\exp\left(is \times \frac{-2\log\Lambda_n + 2\mu_n}{2n\sigma_n}\right)\right\} \qquad \text{(A.30)}\\
&= \left(\frac{2e}{n}\right)^{-npti/2}(1-ti)^{-np(1-ti)/2}\frac{\Gamma_p[\{n(1-ti)-1\}/2]}{\Gamma_p\{(n-1)/2\}}\exp\left(\frac{\mu_n si}{n\sigma_n}\right),
\end{aligned}
$$

where i denotes the imaginary unit and $t = s/(n\sigma_n)$. Then the proof of Theorem 2.2.3 utilizes the following inequality result of the characteristic function.

Lemma A.1.3 (Theorem 1.4.9 (Ushakov, 2011))**.** *Let $G_1(x)$ and $G_0(x)$ be two distribution functions with characteristic functions $\psi_1(s)$ and $\psi_0(s)$, respectively. If $G_0(x)$ has a derivative and $\sup_x G_0'(x) \le a < \infty$, then for any positive T and any $b \ge 1/(2\pi)$,*

$$
\sup_x \left|G_1(x) - G_0(x)\right| \le b\int_{-T}^{T}\left|\frac{\psi_1(s)-\psi_0(s)}{s}\right|ds + \frac{c}{T},
$$

where c is a constant that depends on a and b.

We next prove (2.3) and (2.4) in Theorem 2.2.3 for the chi-squared approximations without and with the Bartlett correction, respectively.

(i) Chi-squared approximation. We prove (2.3) by using Lemma A.1.3 to derive an upper bound of the difference $G_1(x) - G_0(x)$, where we consider

$$
G_1(x) = \Pr\left(\frac{-2\log\Lambda_n + 2\mu_n}{2n\sigma_n} \le x\right), \qquad G_0(x) = \Phi(x);
$$

here $\Phi(x)$ denotes the cumulative distribution function of the standard normal distribution. Then the characteristic function of $G_1(x)$ is $\psi_1(s) =$ (A.30), and the characteristic function of $G_0(x)$ is $\psi_0(s) = \exp(-s^2/2)$. To quantify $\psi_1(s) - \psi_0(s)$, we use the following Lemma A.1.4.

Lemma A.1.4. *When* $s = o(\min\{(n/p)^{1/2}, f^{1/6}\})$,

$$\log \psi_1(s) - \log \psi_0(s) = O\left(\frac{p}{n}\right) s + \left(\frac{1}{p} + \frac{p}{n}\right) O\left(s^2\right) + O\left(\frac{s^3}{\sqrt{f}}\right). \tag{A.31}$$

Proof. Please see Section A.2.3.1 on Page 221. □

By Lemmas A.1.3 and A.1.4, we take $T = \min\{(n/p)^{(1-\delta)/2}, f^{(1-\delta)/6}\}$, where $\delta \in (0,1)$ is a small constant, and then

$$\sup_x \left|G_1(x) - G_0(x)\right| \leq b \int_{-T}^{T} \psi_0(s) \left\{ O\left(\frac{p}{n}\right) + \left(\frac{1}{p} + \frac{p}{n}\right) O\left(s\right) + O\left(\frac{s^2}{\sqrt{f}}\right) \right\} ds + \frac{c}{T}. \tag{A.32}$$

Since $\int_{-T}^{T} \psi_0(s) < \infty$, $\int_{-T}^{T} \psi_0(s)s < \infty$, and $\int_{-T}^{T} \psi_0(s)s^2 < \infty$, by $f = \Theta(p^2)$ and (A.32),

$$\sup_x \left|G_1(x) - G_0(x)\right| = O\left\{ \left(\frac{p}{n}\right)^{(1-\delta)/2} + f^{-(1-\delta)/6} \right\}.$$

Consider $x = \{\chi^2_f(\alpha) + 2\mu_n\}(2n\sigma_n)^{-1}$, and then $G_1(x) - G_0(x)$ gives

$$\Pr\left\{-2\log\Lambda_n \leq \chi^2_f(\alpha)\right\} - \Phi\left\{ \frac{\chi^2_f(\alpha) + 2\mu_n}{2n\sigma_n} \right\} = O\left\{ \left(\frac{p}{n}\right)^{(1-\delta)/2} + f^{-(1-\delta)/6} \right\}. \tag{A.33}$$

Then (2.3) is proved by $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$ and $\Pr\{-2\log\Lambda_n > \chi^2_f(\alpha)\} = 1 - \Pr\{-2\log\Lambda_n \leq \chi^2_f(\alpha)\}$.

(ii) Chi-squared approximation with the Bartlett correction. To prove (2.4), we still use (A.32). Now consider $x = \{\chi_f^2(\alpha) + 2\rho\mu_n\}(2\rho n\sigma_n)^{-1}$, and then $G_1(x) - G_0(x)$ gives

$$\Pr\left\{-2\rho\log\Lambda_n \leq \chi_f^2(\alpha)\right\} - \Phi\left\{\frac{\chi_f^2(\alpha) + 2\rho\mu_n}{2\rho n\sigma_n}\right\} = O\left\{\left(\frac{p}{n}\right)^{(1-\delta)/2} + f^{-(1-\delta)/6}\right\}.$$

Remark A.2. *Although Theorem 2.2.3 is inspired by the limit in* (A.2)*, which was first established in Jiang and Yang (2013), Theorem 2.2.3 differs from the existing results by further characterizing the convergence rate of* (A.2) *by Lemma A.1.4. Particularly, Jiang and Yang (2013) proved* (A.2) *when s is considered fixed and the convergence rate is not examined. On the other hand, Lemma A.1.4 allows s changes with n and p, and the difference between the two characteristic functions is characterized by* (A.31)*. Technically, establishing* (A.31) *requires a careful investigation of the asymptotic expansion of the gamma functions, where the technical details are given in Sections A.2.1 and A.2.3.*

Remark A.3. *Since χ_f^2 can be viewed as a summation over f independent χ_1^2 variables, by applying the central limit theorem, we have $\chi_f^2(\alpha) = \sqrt{2f}z_\alpha + f + O(1)$, where z_α denote the upper α-level quantile of the standard normal distribution. For the problem (III), note that μ_n and σ_n in Theorem 2.2.3 are the same as* (A.3) *and* (A.4)*, respectively. Then by the proof of* (A.7) *in Section A.1.1.1, we have $2n\sigma_n/\sqrt{2f} = 1 + O(p/n)$. Consequently, when $f \to \infty$ and $p/n \to 0$,*

$$\Phi\left\{\frac{\chi_f^2(\alpha) + 2\mu_n}{2n\sigma_n}\right\} = \Phi\left(z_\alpha + \frac{f + 2\mu_n}{2n\sigma_n}\right) + O\left(\frac{1}{\sqrt{f}}\right) + O\left(\frac{p}{n}\right).$$

Moreover, by (A.12)*, $(f + 2\mu_n)/(2n\sigma_n) \sim -p^2/(6n)$ when $p/n \to 0$. Thus $-(f + 2\mu_n)/(2n\sigma_n) = \sqrt{2}\vartheta_1(n,p) + o(p^{1/d_1}n^{-1})$, which is of the order of $p^{1/d_1}n^{-1}$ with $d_1 =$*

$1/2$. *When* $p/n^{d_1} \to 0$, *by* $\alpha = \bar{\Phi}(z_\alpha)$ *and Taylor's series of* $\bar{\Phi}(\cdot)$ *at* z_α,

$$\bar{\Phi}\biggl(z_\alpha + \frac{f + 2\mu_n}{2n\sigma_n}\biggr) - \alpha = \frac{\vartheta_1(n,p)}{\sqrt{\pi}} \exp\biggl(-\frac{z_\alpha^2}{2}\biggr) + o\biggl(\frac{p^{1/d_1}}{n}\biggr),$$

which suggests that the first two terms in the right hand side of (2.3) *are consistent with* (2.1). *Similarly, for the chi-squared approximation with the Bartlett correction, when* $f \to \infty$ *and* $p/n \to 0$,

$$\Phi\biggl\{\frac{\chi_f^2(\alpha) + 2\rho\mu_n}{2\rho n\sigma_n}\biggr\} = \Phi\left(z_\alpha + \frac{f + 2\rho\mu_n}{2\rho n\sigma_n}\right) + O\left(\frac{1}{\sqrt{f}}\right) + O\left(\frac{p}{n}\right).$$

By (A.15), *we have* $-(f + 2\rho\mu_n)/(2\rho n\sigma_n) = \sqrt{2}\vartheta_2(n,p) + o(p^{2/d_2}n^{-2})$, *which is of the order of* $p^{2/d_2}n^{-2}$ *with* $d_2 = 2/3$. *Thus when* $p^{2/d_2}n^{-2} \to 0$, *we also know that the first two terms in the right hand side of* (2.4) *are consistent with* (2.2). *For other likelihood ratio tests (II)–(VI), similar conclusions also hold by the proofs in Section A.1.2.*

In the following of this section, we provide the proofs of other testing problems following similar arguments to that in Section A.1.1. Particularly, for tests (I)–(II) and (IV)–(VII), Theorems 2.2.1 and 2.2.4 are proved in Section A.1.2; Theorems 2.2.2 and 2.2.5 are proved in Section A.1.4, Theorems 2.2.3 and 2.2.6 are proved in Section A.1.5. Proposition 2.2.1 is proved in Section A.1.3.

A.1.2 Proof of Theorems 2.2.1 and 2.2.4

When p is fixed, the chi-squared approximations hold by the classical multivariate analysis (Anderson, 2003; Muirhead, 2009). Therefore, without loss of generality, the proofs below focus on $p \to \infty$. In addition, we note that the analysis of taking subsequences in Section A.1.1.1 can be used similarly in the following proofs, and thus we consider without loss of generality that the sequence p/n has a limit below. We next study six likelihood ratio tests in the following subsections separately.

A.1.2.1 Proof of Theorem 2.2.1 (I): One-Sample Mean Vector Test

Similarly to the proof above, we derive the necessary and sufficient conditions for the chi-squared approximations by examining the moment generating functions. Note that testing one-sample mean vector can be viewed as testing coefficient vector $\boldsymbol{\mu}$ of the multivariate linear regression $\mathbf{x}_i = 1 \times \boldsymbol{\mu} + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Motivated by the approximate expansion of multivariate Gamma function in Jiang and Yang (2013), He et al. (2021a) studied the moment generating function of the likelihood ratio test in high-dimensional multivariate linear regression. Particularly, by Theorem 3 in He et al. (2021a), we know that when $n, p \to \infty$ and $n - p \to \infty$, (A.2) holds with

$$\mu_n = \frac{n}{2} \left\{ (n-p-3/2) \log \frac{(n-p)(n-1)}{n(n-1-p)} + \log\left(1 - \frac{p}{n}\right) + p \log\left(1 - \frac{1}{n}\right) \right\}, \tag{A.34}$$

$$\sigma_n^2 = \frac{1}{2} \left\{ \log\left(1 - \frac{p}{n}\right) - \log\left(1 - \frac{p}{n-1}\right) \right\}. \tag{A.35}$$

Following the analysis in Section A.1.1.1, we know that to derive the necessary and sufficient conditions for the chi-squared approximations without and with the Bartlett correction, it is equivalent to examine (A.7)–(A.8) and (A.9)–(A.10), respectively, with μ_n in (A.34) and σ_n in (A.35).

(I.i) The chi-squared approximation. When $p/n \to 0$, we apply Theorem 1 in He et al. (2021a), and know that (A.7)–(A.8) hold if and only if $p^3/n^2 \to 0$. When $p/n \to C \in (0, 1]$, we have

$$2\sigma_n^2 = \log\left\{ 1 + \left(1 - \frac{p}{n-1}\right)^{-1} \frac{p}{n(n-1)} \right\} \sim \frac{C}{n(1-C)},$$

and then $\sqrt{2f}/(2n\sigma_n) = \sqrt{2p}/(2n\sigma_n) \to \sqrt{1-C} < 1$. Therefore (A.7) fails, which suggests that the classical chi-squared approximation fails.

(I.ii) The chi-squared approximation with the Bartlett correction. When $p/n \to 0$, we

apply Theorem 2 in He et al. (2021a), and know that (A.9)–(A.10) hold if and only if $p^5/n^4 \to 0$. When $p/n \to C \in (0,1]$ and $n-p\to\infty$, we have $\rho \sim 1-C/2$, and then $\sqrt{2f}/(2n\rho\sigma_n) = (1-C/2)^{-1}\sqrt{2p}/(2n\sigma_n) \to (1-C/2)^{-1}\sqrt{1-C} < 1$. Therefore (A.9) fails, which suggests that the classical chi-squared approximation with the Bartlett correction fails.

A.1.2.2 Proof of Theorem 2.2.1 (II): One-Sample Covariance Matrix Test

Similarly to the proof in Section A.1.1.1, by Theorem 1 in Jiang and Yang (2013) and Jiang and Qi (2015), we know that under the conditions of our Theorem 2.2.1 and $p\to\infty$, (A.2) holds with

$$\mu_n = -\frac{(n-1)p}{2} - \frac{n-1}{2}(n-p-3/2)\log\left(1-\frac{p}{n-1}\right), \tag{A.36}$$

$$\sigma_n^2 = -\frac{1}{2}\left\{\frac{p}{n-1}+\log\left(1-\frac{p}{n-1}\right)\right\}\times\frac{(n-1)^2}{n^2}. \tag{A.37}$$

Following the analysis above, we know that to derive the necessary and sufficient conditions for the chi-squared approximations without and with the Bartlett correction, it is equivalent to examine (A.7)–(A.8) and (A.9)–(A.10), respectively, with μ_n in (A.36) and σ_n in (A.37). As analyzed in Section A.1.1.1, it suffices to discuss two cases $\lim_{n\to\infty} p/n = 0$ and $\lim_{n\to\infty} p/n = C \in (0,1]$ below.

(II.i) The chi-squared approximation.

Case (II.i.1) $\lim_{n\to\infty} p/n = 0$. As $\sqrt{2f}\sim p$, and (A.37) and (A.4) are asymptotically the same, by the proof in Section A.1.1.1, we know that (A.7) holds under this case. We next show that (A.8) holds if and only if $p^2/n\to 0$. By (A.7) and $\sqrt{2f}\sim p$, (A.8) is equivalent to $p^{-1}(f+2\mu_n)\to 0$. By Taylor's expansion of μ_n in (A.36), we obtain

$$\mu_n = \frac{-(n-1)p}{2} + \frac{(n-1)}{2}(n-p-3/2)\left\{\frac{p}{n-1}+\frac{p^2}{2(n-1)^2}+\frac{p^3}{3(n-1)^3}+O\left(\frac{p^4}{n^4}\right)\right\}.$$

Through calculations, we obtain

$$p^{-1}(f+2\mu_n) = p^{-1} \times \left\{ -\frac{p^2}{2} + \frac{p^2(n-p)}{2(n-1)} + \frac{p^3 n}{3(n-1)^2} + o(p) + O\left(\frac{p^4}{n^2}\right) \right\}$$
$$= p^{-1} \left\{ -\frac{p^3}{6n} + o(p) + O\left(\frac{p^4}{n^2}\right) \right\} = -\frac{p^2}{6n}\{1+o(1)\} + o(1),$$

which goes to 0 if and only if $p^2/n \to 0$.

Case (II.i.2) $\lim_{n\to\infty} p/n = C \in (0,1]$. Similarly, as (A.37) and (A.4) are asymptotically equal, we can apply the analysis same as Section A.1.1.1, and know that the chi-squared approximation fails under this case.

(II.ii) The chi-squared approximation with the Bartlett correction.

Case (II.ii.1) $\lim_{n\to\infty} p/n = 0$. Under this case, we know (A.9) holds since $\rho = 1 + O(p/n) \to 1$ and $p/(2n\sigma_n) \to 1$ as shown above. Given (A.9), to prove (A.10), it is equivalent to prove $p^{-1}(f + 2\rho\mu_n) \to 0$. By Taylor's expansion of μ_n in (A.37), we have

$$\mu_n = -\frac{p(n-1)}{2} + \frac{(n-p-3/2)(n-1)}{2}$$
$$\times \left\{ \frac{p}{n-1} + \frac{p^2}{2(n-1)^2} + \frac{p^3}{3(n-1)^3} + \frac{p^4}{4(n-1)^4} + O\left(\frac{p^5}{n^5}\right) \right\}.$$

After calculations, we obtain

$$2\rho\mu_n = -p\left(p+\frac{1}{2}\right) + \frac{p^3}{3(n-1)} + \frac{p^2(n-p)}{2(n-1)} - \frac{p^3(n-p)}{6(n-1)^2}$$
$$+ \frac{p^3(n-p)}{3(n-1)^2} - \frac{p^4 n}{9(n-1)^3} + \frac{p^4 n}{4(n-1)^3} + o(p) + O\left(\frac{p^5}{n^3}\right).$$

It follows that $f + 2\rho\mu_n = -p^4 n^{-2}/36 + o(p) + O(p^5 n^{-3})$. Therefore $p^{-1}\{f + \mu_n\rho(n-1)\} \to 0$ if and only if $p^3/n^2 \to 0$.

Case (II.ii.2) $\lim_{n\to\infty} p/n = C \in (0,1]$. Under this case, we have $\rho \to 1 - C/3$. Similarly, as (A.37) (A.4) are asymptotically equal, we can apply the proof same as in Section A.1.1.1, and know that the chi-squared approximation with the Bartlett correction also fails under this case.

A.1.2.3 Proof of Theorem 2.2.4 (IV): Testing the Equality of Several Mean Vectors

Note that testing the equality of several mean vectors can be viewed as testing the coefficient matrix in multivariate linear regression; see, Section 10.7 in Muirhead (2009). Similarly to Section A.1.2.1, by Theorem 3 in He et al. (2021a), we know that when $n, p \to \infty$ and $n - p \to \infty$, (A.2) holds with

$$\mu_n = \frac{n}{2}\Bigg\{(n-p-k-1/2)\log\frac{(n-1-p)(n-k)}{(n-p-k)(n-1)} + (k-1)\log\frac{(n-1-p)}{(n-1)} + p\log\frac{(n-k)}{(n-1)}\Bigg\}, \tag{A.38}$$

$$\sigma_n^2 = \frac{1}{2}\left\{\log\left(1-\frac{p}{n-1}\right) - \log\left(1-\frac{p}{n-k}\right)\right\}. \tag{A.39}$$

Following the analysis in Section A.1.1.1, we know to derive the necessary and sufficient conditions for the chi-squared approximations without and with the Bartlett correction, it is equivalent to examine (A.7)–(A.8) and (A.9)–(A.10), respectively, with μ_n in (A.38) and σ_n in (A.39).

(IV.i) The chi-squared approximation. When $p/n \to 0$, we apply Theorem 1 in He et al. (2021a), and know that (A.7)–(A.8) hold if and only if $p^3/n^2 \to 0$. When $p/n \to C \in (0, 1]$ and $n - p \to \infty$, we have $\sigma_n^2 \sim C(k-1)/\{2n(1-C)\}$, and then $\sqrt{2f}/(2n\sigma_n) = \sqrt{2(k-1)p}/(2n\sigma_n) \to \sqrt{1-C} < 1$. Therefore (A.7) fails, which suggests that the classical chi-squared approximation fails.

(IV.ii) The chi-squared approximation with the Bartlett correction. When $p/n \to 0$, we apply Theorem 2 in He et al. (2021a), and know that (A.9)–(A.10) hold if and only if $p^5/n^4 \to 0$. When $p/n \to C \in (0, 1]$ and $n - p \to \infty$, we have $\rho \sim 1 - C/2$, and then

$\sqrt{2f}/(2n\rho\sigma_n) = (1-C/2)^{-1}\sqrt{2p}/(2n\sigma_n) \to (1-C/2)^{-1}\sqrt{1-C} < 1$. Therefore (A.9) fails, which suggests that the classical chi-squared approximation with the Bartlett correction fails.

A.1.2.4 Proof of Theorem 2.2.4 (V): Testing the Equality of Several Covariance Matrices

Similarly to the proof in Section A.1.1.1, by Theorem 4 in Jiang and Yang (2013) and Jiang and Qi (2015), we know that under the conditions of Theorem 2.2.4 and $p \to \infty$, (A.2) holds with

$$\mu_n = \frac{1}{4}\left\{(n-k)(2n-2p-2k-1)\log\left(1-\frac{p}{n-k}\right)\right. \tag{A.40}$$
$$\left. -\sum_{i=1}^{k}(n_i-1)(2n_i-2p-3)\log\left(1-\frac{p}{n_i-1}\right)\right\},$$
$$\sigma_n^2 = \frac{(n-k)^2}{2n^2}\left\{\log\left(1-\frac{p}{n-k}\right) - \sum_{i=1}^{k}\left(\frac{n_i-1}{n-k}\right)^2\log\left(1-\frac{p}{n_i-1}\right)\right\}. \tag{A.41}$$

Following the analysis in Section A.1.1.1, we next derive the equivalent conditions for (A.7)–(A.8) and (A.9)–(A.10), respectively, with μ_n in (A.40) and σ_n in (A.41).

(V.i) The chi-squared approximation.

Case (V.i.1) $\lim_{n\to\infty} p/n = 0$. Under this case, we show that (A.7) holds. By Taylor's expansion,

$$\begin{aligned}\sigma_n^2 &= \frac{(n-k)^2}{2n^2}\left[-\frac{p}{n-k} - \frac{p^2}{2(n-k)^2}\right.\\ &\qquad \left. + \sum_{i=1}^{k}\left(\frac{n_i-1}{n-k}\right)^2\left\{\frac{p}{n_i-1} + \frac{p^2}{2(n_i-1)^2}\right\} + O\left(\frac{p^3}{n^3}\right)\right]\\ &= \frac{(k-1)p^2}{4n^2}\left\{1+o(1)\right\},\end{aligned}$$

where we use $n_i = \Theta(n)$. As $\sqrt{2f} \sim p\sqrt{k-1}$, we have (A.7) holds. Given (A.7), we know that (A.8) is equivalent to $(2f + 4\mu_n)/(2p\sqrt{k-1}) \to 0$. Through Taylor's expansion, we obtain

$$\begin{aligned}
4\mu_n = & -p(2n-2p-2k-1) - \frac{(n-p)p^2}{n-k} - \frac{2(n-p)p^3}{3(n-k)^2} + o\left(\frac{p^3}{n}\right) + o(p) \\
& + \sum_{i=1}^{k} p(2n_i - 2p - 3) + \sum_{i=1}^{k} \frac{(n_i - p)\, p^2}{(n_i - 1)} + \sum_{i=1}^{k} \frac{2\,(n_i - p)\, p^3}{3\,(n_i - 1)^2} + o\left(\frac{p^3}{n}\right) \\
= & \; p(p - kp - k + 1) + \frac{p^3}{3(n-k)} - \sum_{i=1}^{k} \frac{p^3}{3(n_i - 1)} + o\left(\frac{p^3}{n}\right) + o(p).
\end{aligned}$$

By $f = p(p+1)(k-1)/2$, we have

$$2f + 4\mu_n = \frac{p^3}{3}\left(\frac{1}{n-k} - \sum_{i=1}^{k} \frac{1}{n_i - 1}\right) + o\left(\frac{p^3}{n}\right) + o(p) = \Theta(p^3/n) + o(p), \quad \text{(A.42)}$$

where we use the fact that $(n-k)^{-1} - \sum_{i=1}^{k}(n_i - 1)^{-1} > 0$. It follows that $(2f + 4\mu_n)/(2p\sqrt{k-1}) = \Theta(p^2/n)$, which converges to 0 if and only if $p^2/n \to 0$.

Case (V.i.2) $\lim_{n\to\infty} p/n = C \in (0, 1]$. Under this case, we show that (A.7) and (A.8) do not hold at the same time. Particularly, (A.7) and (A.8) together induce $4(\mu_n + n^2\sigma_n^2)/(2f) \to 0$, which indicates $2(\mu_n + n^2\sigma_n^2)n^{-2} \to 0$, and thus $g_1(C) = 0$, where we define $g_1(C) = (2 - C)\log(1 - C) - \sum_{i=1}^{k} \delta_i(2\delta_i - C)\log(1 - C\delta_i^{-1})$, and we assume $n_i/n \to \delta_i \in (0, 1)$ for $i = 1, \ldots, k$. As $p/n = (p/n_i) \times (n_i/n) < n_i/n$, we have $0 < C \le \delta_i < 1$ for $i = 1, \ldots, k$. We next show that $g_1(C) > 0$ for $C \in (0, \min_{i=1,\ldots,k} \delta_i]$ by taking derivative of $g_1(C)$. Specifically, by $\sum_{i=1}^{k} \delta_i = 1$ and calculations, we have

$$\begin{aligned}
g_1'(C) = & \sum_{i=1}^{k} \delta_i \left\{ -\log(1 - C) - (1 - C)^{-1} + \log(1 - C\delta_i^{-1}) + \delta_i(\delta_i - C)^{-1} \right\}, \\
g_1''(C) = & \sum_{i=1}^{k} \delta_i \times C \left\{ -(1 - C)^{-2} + (\delta_i - C)^{-2} \right\}.
\end{aligned}$$

When $0 < C \le \delta_i < 1$ for $i = 1, \ldots, k$, we have $g_1''(C) > 0$ and thus $g_1'(C)$ is a monotonically increasing function of C. As $g_1'(0) = 0$, $g_1'(C) > 0$ when $0 < C < 1$

and then $g_1(C)$ is also monotonically increasing. By $g_1(0) = 0$, we further obtain $g_1(C) > 0$ when $0 < C < 1$, which contradicts with $g_1(C) = 0$. As a result, we know (A.7) and (A.8) do not hold simultaneously, which suggests that the chi-squared approximation fails.

(V.ii) The chi-squared approximation with the Bartlett correction. When $\lim_{n\to\infty} p/n = 0$, since $\rho = 1 + O(p/n) \to 1$ and (A.7) is proved above, we know (A.9) holds. Given (A.9), as $f \sim p^2(k-1)/2$, to prove (A.10), it is equivalent to show $(2f + 4\rho\mu_n)/p \to 0$, which is also equivalent to $(2f + 4\mu_n - 4\Delta_n\mu_n)/p \to 0$, where we redefine in this subsection that

$$\Delta_n = \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \times \tilde{D}_{n,1}, \qquad \tilde{D}_{n,1} = \sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{n-k}.$$

Similarly to the analysis of (A.42), through Taylor's expansion of μ_n in (A.40), we obtain

$$2f + 4\mu_n = -\frac{p^3}{3} \times \tilde{D}_{n,1} - \frac{p^4}{6} \times \tilde{D}_{n,2} + o\left(\frac{p^4}{n^2}\right) + o(p), \tag{A.43}$$

where $\tilde{D}_{n,2} = \sum_{i=1}^{k}(n_i - 1)^{-2} - (n-k)^{-2}$. Moreover, by (A.42) and $\Delta_n = O(p/n) = o(1)$, we have

$$4\Delta_n\mu_n = \Delta_n\left(-\frac{p^3}{3} \times \tilde{D}_{n,1} - 2f\right) + o\left(\frac{p^4}{n^2}\right) + o(p), \tag{A.44}$$

Combining (A.43) and (A.44), we have

$$\begin{aligned} & 2f + 4\mu_n - 4\Delta_n\mu_n \\ = & -\frac{p^3}{3} \times \tilde{D}_{n,1} - \frac{p^4}{6} \times \tilde{D}_{n,2} + \Delta_n\left(\frac{p^3}{3} \times \tilde{D}_{n,1} + 2f\right) + o\left(\frac{p^4}{n^2}\right) + o(p), \\ = & \frac{p^4}{18(k-1)}\left\{2\tilde{D}_{n,1}^2 - 3(k-1)\tilde{D}_{n,2}\right\} + o\left(\frac{p^4}{n^2}\right) + o(p), \end{aligned} \tag{A.45}$$

where we use $\tilde{D}_{n,1} = \Theta(n^{-1})$, $\tilde{D}_{n,2} = \Theta(n^{-2})$, $\Delta_n = p\tilde{D}_{n,1}/\{3(k-1)\} + o(p/n)$, and $2\Delta_n f = p^3\tilde{D}_{n,1}/3 + o(p)$.

We next show that (A.45) $= \Theta(p^4 n^{-2})$. In particular, in this subsection, we redefine $\delta_i = (n_i - 1)/(n-k)$, which satisfies $\sum_{i=1}^k \delta_i = 1$. Then by the definitions of $\tilde{D}_{n,1}$ and $\tilde{D}_{n,2}$, we calculate that

$$\begin{aligned}
&(n-k)^2 \times \{2\tilde{D}_{n,1}^2 - 3(k-1)\tilde{D}_{n,2}\} \\
&= (5-3k)\sum_{i=1}^k \delta_i^{-2} + 2\sum_{1\le i\neq j\le k} \delta_i^{-1}\delta_j^{-1} - 4\sum_{i=1}^k \delta_i^{-1} + 3k - 1. \qquad \text{(A.46)}
\end{aligned}$$

As $2\delta_i^{-1}\delta_j^{-1} \le \delta_i^{-2} + \delta_j^{-2}$, we have

$$\begin{aligned}
\text{(A.46)} &\le (3-k)\sum_{i=1}^k \delta_i^{-2} - 4\sum_{i=1}^k \delta_i^{-1} + 3k - 1 \\
&\le (3-k)\sum_{i=1}^k \delta_i^{-2} - 4k^2 + 3k - 1, \qquad \text{(A.47)}
\end{aligned}$$

where in the last inequality, we use $\sum_{i=1}^k \delta_i^{-1} \ge k^2(\sum_{i=1}^k \delta_i)^{-1} = k^2$. Therefore (A.47) < 0 when $k \ge 3$. When $k = 2$, as $\delta_1 + \delta_2 = 1$, we have $\delta_1^{-1} + \delta_2^{-1} = \delta_1^{-1}\delta_2^{-1}$ and (A.46) $= -\sum_{i=1}^2 \delta_i^{-2} - 2\sum_{i=1}^2 \delta_i^{-1} + 5$. As $\sum_{i=1}^2 \delta_i^{-1} \ge 2^2$, (A.46) $< -2 \times 2^2 + 5 < 0$. In summary, we know (A.46) < 0 for $k \ge 2$, and thus (A.45) $= \Theta(p^4 n^{-2})$. If follows that $(2f + 4\rho\mu_n)/p \to 0$ if and only if $p^3/n^2 \to 0$. In summary, we know for testing problem (V), the chi-squared approximation with the Bartlett correction works if and only if $p^3/n^2 \to 0$.

A.1.2.5 Proof of Theorem 2.2.4 (VI): Joint Testing the Equality of Several Mean Vectors and Covariance Matrices

Similarly to the proof in Section A.1.1.1, by Theorem 3 in Jiang and Yang (2013) and Jiang and Qi (2015), we know that under the conditions of Theorem 2.2.4 and

$p \to \infty$, (A.2) holds with

$$\mu_n = \frac{1}{4}\left\{-2kp - \sum_{i=1}^{k}\frac{p}{n_i} - nL_{n,p}(2p-2n+3) + \sum_{i=1}^{k} n_i L_{n_i-1,p}(2p-2n_i+3)\right\}, \tag{A.48}$$

$$\sigma_n^2 = \frac{1}{2}\left(L_{n,p} - \sum_{i=1}^{k}\frac{n_i^2}{n^2} \times L_{n_i-1,p}\right), \tag{A.49}$$

where $L_{n,p} = \log(1-p/n)$. Following Section A.1.1.1, we next derive the equivalent conditions for (A.7)–(A.8) and (A.9)–(A.10), respectively, with μ_n in (A.48) and σ_n in (A.49).

(VI.i) The chi-squared approximation.

Case (VI.i.1) $\lim_{n\to\infty} p/n = 0$. Under this case, we show that (A.7) holds. As $-\log(1-x) = x + x^2/2 + O(x^3)$ and $n_i = \Theta(n)$, we obtain

$$\begin{aligned} 2\sigma_n^2 = & \sum_{i=1}^{k}\frac{n_i^2}{n^2}\left\{\frac{p}{n_i-1} + \frac{p^2}{2(n_i-1)^2}\right\} - \frac{p}{n} - \frac{p^2}{2n^2} + O\left(\frac{p^3}{n^3}\right) \\ = & \frac{kp}{n^2} + \frac{(k-1)p^2}{2n^2} + O\left(\frac{p^3}{n^3}\right), \end{aligned}$$

where in the second equation, we use $(n_i-1)^{-1} = n_i^{-1} + n_i^{-2} + O(n_i^{-3})$ and $(n_i-1)^{-2} = n_i^{-2} + O(n_i^{-3})$. It follows that $2n\sigma_n \sim p\sqrt{k-1}$. By $\sqrt{2f} \sim p\sqrt{k-1}$, we have (A.7). Given (A.7), we know that (A.8) is equivalent to $(2f + 4\mu_n)/p \to 0$. As $p/n = o(1)$, through Taylor's expansion, we obtain

$$\begin{aligned} -n(2p-2n+3)L_{n,p} &= n(2p-2n+3)\left\{\frac{p}{n} + \frac{p^2}{2n^2} + \frac{p^3}{3n^3} + O\left(\frac{p^4}{n^4}\right)\right\} \\ &= p\left\{p + \frac{p^2}{3n} - 2n + 3 + O\left(\frac{p^3}{n^2}\right) + o(1)\right\}. \end{aligned} \tag{A.50}$$

Similarly, by Taylor's expansion and $n_i = \Theta(n)$, we have

$$-n_i(2p - 2n_i + 3)L_{n_i-1,p} = p\left\{p + \frac{p^2}{3n_i} - 2n_i + 3 - 2 + O\left(\frac{p^3}{n_i^2}\right) + o(1)\right\}, \quad \text{(A.51)}$$

where we use $(n_i - 1)^{-1} = n_i^{-1} + n_i^{-2} + O(n_i^{-3})$ and $(n_i - 1)^{-a} = n_i^{-a} + O(n_i^{-3})$ for integers $a \geq 2$. Combining (A.50) and (A.51), we obtain

$$2f + 4\mu_n = \frac{p^3}{3}\Big(\frac{1}{n} - \sum_{i=1}^{k}\frac{1}{n_i}\Big) + O\Big(\frac{p^4}{n^2}\Big) + o(p). \quad \text{(A.52)}$$

As $n^{-1} - \sum_{i=1}^{k} n_i^{-1} = \Theta(n^{-1})$, we have $2f + 4\mu_n = \Theta(p^3 n^{-1})$. Therefore we know $(2f + 4\mu_n)/p \to 0$ if and only if $p^2/n \to 0$.

Case (VI.i.2) $\lim_{n\to\infty} p/n = C \in (0, 1]$. In this subsection, we redefine $\delta_i = n_i/n \in (0, 1)$. Then

$$\frac{4n^2\sigma_n^2}{2f} \to \frac{2}{C^2(k-1)} \times \Big\{\log(1-C) - \sum_{i=1}^{k}\delta_i^2\log(1 - C\delta_i^{-1})\Big\},$$

where $0 < C \leq \delta_i < 1$. Therefore (A.7) induces $g_2(C) = 0$, where we define

$$g_2(C) = \log(1-C) - \sum_{i=1}^{k}\delta_i^2\log(1 - C\delta_i^{-1}) - (k-1)C^2/2.$$

By taking derivative of $g_2(C)$, we obtain $g_2'(0) = 0$, $g_2''(0) = 0$, and

$$g_2'''(C) = \frac{2}{(C-1)^2} - \sum_{i=1}^{k}\frac{2\delta_i^2}{(C-\delta_i)^3} = \sum_{i=1}^{k}\frac{2\delta_i(1-\delta_i)(C^3 - 3\delta_i C + \delta_i^2 + \delta_i)}{(1-C)^3(\delta_i - C)^3}.$$

As $C^3 - 3\delta_i C + \delta_i^2 + \delta_i$ is a monotonically decreasing function of C when $0 < C \leq \delta_i < 1$, and it equals $\delta_i(\delta_i - 1)^2 > 0$ when $C = \delta_i$, we have $g_2'''(C) > 0$ for $0 < C \leq \delta_i$. It follows that $g_2(C)$ is a monotonically increasing function when $0 < C \leq \delta_i < 1$. As $g_2(0) = 0$, we have $g_2(C) > 0$, which contradicts with $g_2(C) = 0$. Therefore, we know that (A.7) does not hold under this case, which implies that the chi-squared approximation fails.

(VI.ii) The chi-squared approximation with the Bartlett correction. When $\lim_{n\to\infty} p/n = 0$, since $\rho = 1 + O(p/n) \to 1$ and (A.7) is proved above, we know (A.9) holds. Given (A.9), as $f \sim p^2(k-1)/2$, to prove (A.10), it is equivalent to show $(2f+4\rho\mu_n)/p \to 0$, which is equivalent to $(2f+4\mu_n-4\Delta_n\mu_n)/p \to 0$, where in this subsection, we redefine

$$\Delta_n = \frac{2p^2+9p+11}{6(p+3)(k-1)} \times D_{n,1}, \qquad D_{n,1} = \sum_{i=1}^{k} \frac{1}{n_i} - \frac{1}{n}.$$

Similarly to (A.50), through Taylor's expansion, we further have

$$n(2p-2n+3)r_n^2 = p\left\{p - 2n + 3 + \frac{p^2}{3n} + \frac{p^3}{6n^2} + O\Big(\frac{p^4}{n^3}\Big) + o(1)\right\}.$$

In addition, similarly to (A.51), we have

$$n_i(2p-2n_i+3)r_{n_i'}^2 = p\left\{p - 2n_i + 3 - 2 + \frac{p^2}{3n_i} + \frac{p^3}{6n_i^2} + O\Big(\frac{p^4}{n_i^3}\Big) + o(1)\right\}. \quad \text{(A.53)}$$

It follows that

$$2f + 4\mu_n = -\frac{p^3}{3}D_{n,1} - \frac{p^4}{6}D_{n,2} + O\Big(\frac{p^5}{n^3}\Big) + o(p), \quad \text{(A.54)}$$

where $D_{n,2} = \sum_{i=1}^{k} n_i^{-2} - n^{-2}$. Moreover, by (A.52) and $\Delta_n = O(p/n) = o(1)$,

$$4\Delta_n\mu_n = \Delta_n\Big(-\frac{p^3}{3}D_{n,1} - 2f\Big) + O\Big(\frac{p^5}{n^3}\Big) + o(p). \quad \text{(A.55)}$$

Combining (A.54) and (A.55), we obtain

$$2f + 4\mu_n - 4\Delta_n\mu_n = \frac{p^4}{18(k-1)}\{2D_{n,1}^2 - 3(k-1)D_{n,2}\} + O\Big(\frac{p^5}{n^3}\Big) + o(p), \quad \text{(A.56)}$$

where we use $D_{n,1} = \Theta(n^{-1})$, $D_{n,2} = \Theta(n^{-2})$, $\Delta_n = pD_{n,1}/\{3(k-1)\} + o(p/n)$, and $2\Delta_n f = p^3 D_{n,1}/3 + o(p)$. Following the analysis of (A.46), we know (A.56) $= \Theta(p^4 n^{-2})$. Therefore, $(2f + \rho\mu_n)/p \to 0$ if and only if $p^3/n^2 \to 0$, which suggests that the chi-squared approximation with the Bartlett correction holds if and only if $p^3/n^2 \to 0$.

A.1.3 Proof of Proposition 2.2.1

This section proves Proposition 2.2.1 following similar arguments to that in Section A.1.1.1. In particular, consider without loss of generality that $p \to \infty$ and p/n has a limit. Following the analysis in Section A.1.2.3, we know that when $n, p \to \infty$, $n - k \to \infty$, and $n - p \to \infty$, (A.2) holds with μ_n in (A.38) and σ_n^2 (A.39). Moreover, to derive the necessary and sufficient conditions for the chi-squared approximations without and with the Bartlett correction, it is equivalent to examine (A.7)–(A.8) and (A.9)–(A.10), respectively, with μ_n in (A.38) and σ_n in (A.39).

(i) The chi-squared approximation. (i.1) When $p/n \to 0$ and $k/n \to 0$, we apply Theorem 1 in He et al. (2021a), and know that (A.7)–(A.8) hold if and only if $\sqrt{pk}(p+k)/n \to 0$. (i.2) When $p/n \to C \in (0, 1]$ and $k/n \to 0$, we have $f \sim C(k-1)n$ and $2\sigma_n^2 \sim C(k-1)/\{n(1-C)\}$. It follows that $\sqrt{2f}/(2n\sigma_n) \sim \sqrt{1-C} < 1$. Thus (A.7) fails, which suggests that the chi-squared approximation fails. (i.3) When $p/n \to 0$ and $k/n \to C \in (0, 1]$, by applying the symmetric substitution technique in Section 10.4 of Muirhead (2009), we can switch k and p and analyze similarly as in the case (i.2) above. Therefore we know the chi-squared approximation also fails here. (i.4) When $p/n \to C_1 \in (0, 1]$ and $k/n \to C_2 \in (0, 1]$, we know $0 < C_1 + C_2 \leq 1$ as $p + k < n$. By the constraint, it then suffices to consider $C_1, C_2 \in (0, 1)$. Note that $2\sigma_n^2 \sim \log\{(1-C_1)(1-C_2)\} - \log(1-C_1-C_2)$ and $2f/n^2 \sim 2C_1C_2$. Thus (A.7) induces $g_4(C_1, C_2) = 0$ where $g_4(C_1, C_2) = C_1C_2 - \log\{(1-C_1)(1-C_2)\} + \log(1-C_1-C_2)$. If $C_1 + C_2 = 1$, $g_4(C_1, C_2) \to -\infty$. We next consider $0 < C_1 + C_2 < 1$. By calculations,

we have $g_4(0, C_2) = 0$, and

$$\frac{\mathrm{d}}{\mathrm{d}C_1} g_4(C_1, C_2) = \frac{C_2\{(C_1-1)(C_1+C_2) - C_1\}}{(1-C_1)(1-C_1-C_2)} < 0,$$

where we use $C_1, C_2 \in (0,1)$ and $0 < C_1 + C_2 < 1$. Similarly to the previous analyses, we know that $g_4(C_1, C_2)$ is monotonically decreasing for $C_1 \in (0,1)$ and thus $g_4(C_1, C_2) < 0$, as $C_1 \in (0,1)$ and $g_4(0, C_2) = 0$. Therefore (A.7) fails, which suggests that the classical chi-squared approximation fails.

(ii) The chi-squared approximation with the Bartlett correction. (ii.1) When $p/n \to 0$ and $k/n \to 0$, we apply Theorem 2 in He et al. (2021a), and know that (A.9)–(A.10) hold if and only if $\sqrt{pk}(p^2 + k^2)/n^2 \to 0$. (ii.2) When $p/n \to C \in (0,1]$ and $k/n \to 0$, we have $\rho \sim 1 - C/2$, and the proof of part (IV.ii) in Section A.1.2.3 can be applied here similarly. Thus the chi-squared approximation fails. (ii.3) When $p/n \to 0$ and $k/n \to C \in (0,1]$, we know the chi-squared approximation also fails by switching k and p symmetrically as in the case (i.3) above. (ii.4) When $p/n \to C_1 \in (0,1]$ and $k/n \to C_2 \in (0,1]$, we know $0 < C_1 + C_2 \leq 1$ as $p + k < n$. Similarly to the case (i.4) above, we consider $C_1, C_2 \in (0,1)$ and $C_1 + C_2 < 1$. Here $\rho \sim 1 - (C_1 + C_2)/2$ and then (A.9) induces $g_5(C_1, C_2) = 0$, where $g_5(C_1, C_2) = 2C_1C_2 - (2 - C_1 - C_2)[\log\{(1 - C_1)(1 - C_2)\} - \log(1 - C_1 - C_2)]$. By calculations, we have $g_5(0, C_2) = 0$, and

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}C_1} g_5(C_1, C_2)|_{C_1=0} = -C_2/(1-C_2) < 0,\\
&\frac{\mathrm{d}^2}{\mathrm{d}^2C_1} g_5(C_1, C_2) = -\frac{C_2\{(C_1+C_2)(C_2-2)+2\}}{(1-C_1)^2(1-C_1-C_2)^2} < 0,
\end{aligned}$$

where we use $(C_1 + C_2)(C_2 - 2) + 2 > 0$ as $0 < C_1 + C_2 < 1$ and $-2 < C_2 - 2 < -1$. Similarly to the analysis above, we know that $g_5(C_1, C_2) < 0$ and thus (A.9) fails, which suggests that the chi-squared approximation with the Bartlett correction fails.

A.1.4 Proofs of Theorems 2.2.2 and 2.2.5

In this section, we prove the results for other testing problems in Theorems 2.2.2 and 2.2.5 following similar analysis to that in Section A.1.1.2. Particularly, for each test, we consider the characteristic function of $-2\eta \log \Lambda_n$ when $\eta = 1$ and ρ; here ρ denotes the corresponding Bartlett correction factor of each test.

By Eq. (20)–(23) in Section 8.2.4 of Muirhead (2009), we know that for the testing problems (I)–(II) and (IV)–(VII), the characteristic functions of the likelihood ratio test statistics take the following general form:

$$\log \mathrm{E}\{\exp(-2it\eta \log \Lambda_n)\} = \varphi(t) - \varphi(0), \tag{A.57}$$

where

$$\begin{aligned}
\varphi(t) = {} & 2it\eta \left(\sum_{k=1}^{K_1} \xi_{1,k} \log \xi_{1,k} - \sum_{j=1}^{K_2} \xi_{2,j} \log \xi_{2,j} \right) \\
& + \sum_{k=1}^{K_1} \log \Gamma\big\{\eta \xi_{1,k}(1-2it) + \tau_{1,k} + \upsilon_{1,k}\big\} \\
& - \sum_{j=1}^{K_2} \log \Gamma\big\{\eta \xi_{2,j}(1-2it) + \tau_{2,j} + \upsilon_{2,j}\big\},
\end{aligned}$$

i denotes the imaginary unit, $\tau_{1,k} = (1-\eta)\xi_{1,k}$, and $\tau_{2,j} = (1-\eta)\xi_{2,j}$. We next consider $\eta = 1$ and ρ for the chi-squared approximation without and with the Bartlett correction, respectively. The values of ρ, K_1, K_2, $\xi_{1,k}$, $\xi_{2,j}$, $\upsilon_{1,k}$, and $\upsilon_{2,j}$ depend on the testing problem, and thus take different values in the following subsections. Moreover, by Muirhead (2009), in each problem, we have $\sum_{k=1}^{K_1} \xi_{1,k} = \sum_{j=1}^{K_2} \xi_{2,k}$, the degrees of freedom f is

$$f = -2 \left\{ \sum_{k=1}^{K_1} \upsilon_{1,k} - \sum_{j=1}^{K_2} \upsilon_{2,j} - \frac{1}{2}(K_1 - K_2) \right\}, \tag{A.58}$$

and the Bartlett correction ρ takes the value

$$\rho = 1 - \frac{1}{f}\left\{\sum_{k=1}^{K_1}\frac{\upsilon_{1,k}^2 - \upsilon_{1,k} + \frac{1}{6}}{\xi_{1,k}} - \sum_{j=1}^{K_2}\frac{\upsilon_{2,j}^2 - \upsilon_{2,j} + \frac{1}{6}}{\xi_{2,j}}\right\}. \tag{A.59}$$

In the following proofs, we use Lemma A.1.5 below to obtain an asymptotic expansion of each characteristic function.

Lemma A.1.5. *For a finite integer L, when $\eta = 1$ or ρ, $p/n \to 0$, and $R_{n,L}$ (in* (A.60) *below) converges to* 0*,*

$$\log \mathrm{E}\{\exp(-2it\eta \log \Lambda_n)\} = -\frac{f}{2}\log(1-2it) + \sum_{l=1}^{L-1}\varsigma_l\{(1-2it)^{-l} - 1\} + R_{n,L},$$

where

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)}\left\{\sum_{k=1}^{K_1}\frac{B_{l+1}(\tau_{1,k}+\upsilon_{1,k})}{(\eta\times\xi_{1,k})^l} - \sum_{j=1}^{K_2}\frac{B_{l+1}(\tau_{2,j}+\upsilon_{2,j})}{(\eta\times\xi_{2,j})^l}\right\},$$

$B_{l+1}(\cdot)$ denotes the $(l+1)$-th Bernoulli polynomial; see, e.g., Eq. (25) in Section 8.2.4 of Muirhead (2009), and $R_{n,L}$ denotes the remainder which is of the order of

$$R_{n,L} = O\left(\sum_{k=1}^{K_1}\frac{|\tau_{1,k}+\upsilon_{1,k}|^{L+1}}{|\eta\xi_{1,k}|^L} + \sum_{j=1}^{K_2}\frac{|\tau_{2,j}+\upsilon_{2,j}|^{L+1}}{|\eta\xi_{2,j}|^L}\right). \tag{A.60}$$

Proof. Please see Section A.2.2.16 on Page 220. □

We next examine each testing problem based on Lemma A.1.5.

A.1.4.1 Proof of Theorem 2.2.2 (I): One-Sample Mean Test

Recall that in Section A.1.2.1, we mention that the one-sample mean test can be viewed as testing the coefficient vector of a multivariate linear regression model. By Section 10.5 in Muirhead (2009), we know that in this problem, $K_1 = 1$, $K_2 = 1$,

$\xi_{1,1} = n/2$, $\xi_{2,1} = n/2$, $\upsilon_{1,1} = -p/2$, $\upsilon_{2,1} = 0$, $f = p$ and $\rho = 1 - (p/2+1)/n$. We next discuss the chi-squared approximation without and with the Bartlett correction, respectively.

(i) Chi-squared approximation. Consider $\rho = 1$ and $p^3/n^2 \to 0$. Then $\tau_{1,1} = \tau_{2,1} = 0$,

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)} \times \frac{1}{(n/2)^l} \left\{ B_{l+1}\left(-\frac{p}{2}\right) - B_{l+1}(0) \right\}, \tag{A.61}$$

and for any finite integer L, $R_{n,L} = O(p^{L+1}n^{-L})$. Since $B_{l+1}(\cdot)$ is a polynomial of order $l+1$, then $\varsigma_l = O(p^{l+1}/n^l)$. By Lemma A.1.5, when $p^3/n^2 \to 0$, $R_{n,3} = O(p^4 n^{-3}) \to 0$, and

$$\begin{aligned} \mathrm{E}\{\exp(-2it\log\Lambda_n)\} &= (1-2it)^{-\frac{f}{2}} \prod_{l=1}^{2} \exp\left[\varsigma_l\big\{(1-2it)^{-l} - 1\big\}\right]\big\{1 + O(p^4 n^{-3})\big\} \\ &= (1-2it)^{-\frac{f}{2}}\big\{1 + V_1(t) + V_2(t) + V_1(t)V_2(t)\big\}\big\{1 + O(p^4 n^{-3})\big\}, \end{aligned}$$

where $V_l(t)$ is defined as in (A.18) on Page 144. Then similarly to the proof in Section A.1.1.2, by the inversion property of the characteristic function, we obtain

$$\begin{aligned} &\Pr(-2\log\Lambda_n \le x) \qquad\qquad (\text{A.62}) \\ =& \Bigg\{ \Pr(\chi^2_f \le x) + \sum_{v=1}^{\infty} \frac{\varsigma_1^v}{v!} \sum_{w=0}^{v} \binom{v}{w} \Pr(\chi^2_{f+2w} \le x)(-1)^{v-w} \\ &+ \sum_{v=1}^{\infty} \frac{\varsigma_2^v}{v!} \sum_{w=0}^{v} \binom{v}{w} \Pr(\chi^2_{f+4w} \le x)(-1)^{v-w} \\ &+ \sum_{\substack{v_1\ge 1;\ 0\le w_1 \le v_1 \\ v_2 \ge 1;\ 0 \le w_2 \le v_2}} \frac{\varsigma_1^{v_1}\varsigma_2^{v_2}}{v_1! v_2!} \binom{v_1}{w_1}\binom{v_2}{w_2} \Pr(\chi^2_{2f+2w_1+4w_2} \le x)(-1)^{v_1-w_1+v_2-w_2} \Bigg\} \\ &\times \left\{ 1 + O\left(\frac{p^4}{n^3}\right) \right\}. \end{aligned}$$

When $x = \chi^2_f(\alpha)$, by Propositions A.1.1 and A.1.2, and $\varsigma_l = O(p^{l+1}/n^l)$, we have

$$\Pr(-2\log\Lambda_n \le x) = \Pr(\chi^2_f \le x) + \varsigma_1\big\{\Pr(\chi^2_{f+2} \le x) - \Pr(\chi^2_f \le x)\big\} + o(p^{3/2}/n).$$

Particularly, by Lemma A.1.2,

$$\Pr(\chi^2_{f+2} \le x) - \Pr(\chi^2_f \le x) = -\frac{1}{\sqrt{f\pi}}\exp\left(-\frac{z_\alpha^2}{2}\right)\Big\{1 + O(f^{-1/2})\Big\},$$

and we compute $\varsigma_1 = (p^2+2p)/(4n)$. In Theorem 2.2.2, we have $\vartheta_1(n,p) = \varsigma_1/\sqrt{f}$.

(ii) Chi-squared approximation with the Bartlett correction. By choosing the Bartlett correction factor ρ as in (A.59), we have $\varsigma_1 = 0$; see, e.g., Section 8.2.4 in Muirhead (2009). Specifically, in this problem, $\rho = 1 - (p+2)/(2n)$, $\rho\xi_{1,1} = \rho\xi_{2,1} = n/2 - (p+2)/4$, $\tau_{1,1} = \tau_{2,1} = (p+2)/4$, $\upsilon_{1,1} = -p/2$, $\upsilon_{2,1} = 0$, and then

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)(\rho\times\xi_{1,1})^l}\left\{B_{l+1}\left(-\frac{p-2}{4}\right) - B_{l+1}\left(\frac{p+2}{4}\right)\right\}.$$

We calculate $\varsigma_2 = p(p^2-4)\{48(\rho n)^2\}^{-1}$, $\varsigma_3 = 0$, and $\varsigma_l = O(p^{l+1}n^{-l})$ for $l \ge 4$. Similarly to the proof in Section A.1.1.2, when $p^5/n^4 \to 0$, we have

$$\begin{aligned}
&\mathrm{E}\{\exp(-2it\rho\log\Lambda_n)\} \\
&= (1-2it)^{-\frac{f}{2}}\big\{1 + V_2(t) + V_4(t) + V_2(t)V_4(t)\big\}\big\{1 + O(p^6/n^5)\big\},
\end{aligned}$$

and thus

$$
\begin{aligned}
&\Pr(-2\rho\log\Lambda_n \le x) \qquad\qquad (\text{A.63})\\
&= \Bigg\{ \Pr(\chi^2_f \le x) + \sum_{v=1}^{\infty} \frac{\varsigma_2^v}{v!} \sum_{w=0}^{v} \binom{v}{w} \Pr(\chi^2_{f+4w} \le x)(-1)^{v-w}\\
&\quad + \sum_{v=1}^{\infty} \frac{\varsigma_4^v}{v!} \sum_{w=0}^{v} \binom{v}{w} \Pr(\chi^2_{f+8w} \le x)(-1)^{v-w}\\
&\quad + \sum_{\substack{v_2\ge 1;\ 0\le w_2\le v_2\\ v_4\ge 1;\ 0\le w_4\le v_4}} \frac{\varsigma_2^{v_2}\varsigma_4^{v_4}}{v_2!v_4!}\binom{v_2}{w_2}\binom{v_4}{w_4} \Pr(\chi^2_{2f+4w_2+8w_4} \le x)(-1)^{v_2-w_2+v_4-w_4} \Bigg\}\\
&\quad \times\Big\{1+O\Big(\frac{p^6}{n^5}\Big)\Big\}.
\end{aligned}
$$

Note that $\varsigma_2 = \Theta(p^3n^{-2})$ and $\varsigma_4 = \Theta(p^5n^{-4})$. By applying proposition A.1.1 with $h = 2$,

$$
\sum_{v=1}^{\infty} \frac{\varsigma_2^v}{v!} \sum_{w=0}^{v} \binom{v}{w} \Pr(\chi^2_{f+4w} \le x)(-1)^{v-w} = \sum_{v=1}^{\infty} \Big\{O\big(\varsigma_2 p^{-1/2}\big)\Big\}^v = \Theta(p^{5/2}n^{-2}).
$$

By applying proposition A.1.1 with $h = 4$, we have

$$
\begin{aligned}
\sum_{v=1}^{\infty} \frac{\varsigma_4^v}{v!} \sum_{w=0}^{v} \binom{v}{w} \Pr(\chi^2_{f+8w} \le x)(-1)^{v-w} &= \sum_{v=1}^{\infty} \Big\{O\big(\varsigma_4 p^{-1/2}\big)\Big\}^v\\
&= O\big(p^{9/2}n^{-4}\big) = o(p^{5/2}n^{-2}),
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{\substack{v_2\ge 1;\ 0\le w_2\le v_2\\ v_4\ge 1;\ 0\le w_4\le v_4}} \frac{\varsigma_2^{v_2}\varsigma_4^{v_4}}{v_2!v_4!}\binom{v_2}{w_2}\binom{v_4}{w_4} \Pr(\chi^2_{2f+4w_2+8w_4} \le x)(-1)^{v_2-w_2+v_4-w_4}\\
&= \sum_{v_2\ge 1} \big\{O\big(\varsigma_2 p^{-1/2}\big)\big\}^{v_2} \sum_{v_4\ge 1} \frac{\{O(\varsigma_4)\}^{v_4}}{v_4!} = o\big(p^{5/2}n^{-2}\big).
\end{aligned}
$$

In summary, by (A.63),

$$\Pr(-2\rho\log\Lambda_n \le x) = \Pr(\chi_f^2 \le x) + \varsigma_2\big\{\Pr(\chi_{f+4}^2 \le x) - \Pr(\chi_f^2 \le x)\big\} + o\big(p^{5/2}n^{-2}\big).$$

Particularly, by Lemma A.1.2,

$$\Pr(\chi_{f+4}^2 \le x) - \Pr(\chi_f^2 \le x) = -\frac{2}{\sqrt{f\pi}}\exp\left(-\frac{z_\alpha^2}{2}\right)\Big\{1 + O(f^{-1/2})\Big\}.$$

In Theorem 2.2.2 (I), $\vartheta_2(n,p) = 2\varsigma_2/\sqrt{f}$.

A.1.4.2 Proof of Theorem 2.2.2 (II): Testing One-Sample Covariance Matrix

In this problem, by Section 8.3.3 in Muirhead (2009), we know $f = (p+2)(p-1)/2$, and

- $K_1 = p$, $K_2 = 1$;
- $\xi_{1,k} = (n-1)/2$, $\upsilon_{1,k} = -(k-1)/2$ for $k = 1,\dots,K_1$;
- $\xi_{2,1} = p(n-1)/2$, $\upsilon_{2,1} = 0$.

(i) Chi-squared approximation. Consider $\rho = 1$ and $p^2/n \to 0$. Then $\tau_{1,k} = 0$ for $k = 1,\dots,K_1$, $\tau_{2,1} = 0$, and

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)}\left\{\sum_{k=1}^{p}\left(\frac{2}{n-1}\right)^l B_{l+1}\left(-\frac{k-1}{2}\right) - \frac{2}{p(n-1)}B_{l+1}(0)\right\},$$

which satisfies $\varsigma_l = O(p^{l+2}/n^l)$. By Lemma A.1.5,

$$\mathrm{E}\{\exp(-2it\log\Lambda_n)\} = (1-2it)^{-\frac{f}{2}}\big\{1 + V_1(t) + V_2(t) + V_1(t)V_2(t)\big\}\big\{1 + O(p^5/n^3)\big\},$$

where $V_l(t)$ is defined as in (A.18). Similarly to Section A.1.1.2, by the inversion property of the characteristic functions, and Propositions A.1.1 and A.1.2, we obtain (A.22). We calculate

$$\begin{aligned}\varsigma_1 &= \frac{1}{2}\left[\sum_{k=1}^{p}\frac{2}{n-1}\left\{\left(-\frac{k-1}{2}\right)^2-\left(-\frac{k-1}{2}\right)+\frac{1}{6}\right\}-\frac{2}{p(n-1)}\times\frac{1}{6}\right]\\ &= \frac{2p^3+3p^2-p-4/p}{24(n-1)}.\end{aligned}$$

The conclusion then follows by Lemma A.1.2 and $\vartheta_1(n,p)=\varsigma_1/\sqrt{f}$.

(ii) Chi-squared approximation with the Bartlett correction. In this problem, consider

$$\rho = 1-\frac{2p^2+p+2}{6p(n-1)},$$

and $p^3/n^2 \to 0$. Then $\tau_{1,k}=(2p^2+p+2)/(12p)$ for $k=1,\ldots,p$, and $\tau_{2,1}=(2p^2+p+2)/12$. It follows that

$$\begin{aligned}\varsigma_l = &\frac{(-1)^{l+1}}{l(l+1)}\left\{\frac{\rho(n-1)}{2}\right\}^{-l}\\ &\times\left\{\sum_{k=1}^{p}B_{l+1}\left(\frac{2p^2+p+2}{12p}-\frac{k-1}{2}\right)-p^{-l}B_{l+1}\left(\frac{2p^2+p+2}{12}\right)\right\}.\end{aligned}$$

In particular, we calculate

$$\varsigma_2=\frac{(p-2)(p-1)(p+2)}{288p^2\rho^2(n-1)^2}(2p^3+6p^2+3p+2).$$

Similarly to Section A.1.1.2, by the inversion property of the characteristic functions, and Propositions A.1.1 and A.1.2, we obtain (A.29). The conclusion then follows by Lemma A.1.2 and $\vartheta_2(n,p)=2\varsigma_2/\sqrt{f}$.

A.1.4.3 Proof of Theorem 2.2.5 (IV): Testing the Equality of Several Mean Vectors

Recall that in Section A.1.2.3, we show that this testing problem can be viewed as testing the coefficient matrix in multivariate linear regression. Then by Eq. (3) in Section 10.5.3 in Muirhead (2009), we know that in this problem, $f = (k-1)p$, and

- $K_1 = k-1$, $K_2 = k-1$;
- $\xi_{1,j_1} = n/2$, $\upsilon_{1,j_1} = -(j_1+p)/2$, $\quad j_1 = 1, \dots, k-1$;
- $\xi_{2,j_2} = n/2$, $\upsilon_{2,j_2} = -j_2/2$, $\quad j_2 = 1, \dots, k-1$.

(i) Chi-squared approximation. It follows that

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)} \left(\frac{2}{n}\right)^l \left\{ \sum_{j_1=1}^{k-1} B_{l+1}\left(-\frac{j_1+p}{2}\right) - \sum_{j_2=1}^{k-1} B_{l+1}\left(-\frac{j_2}{2}\right) \right\},$$

which is $O(p^{l+1}n^{-l})$ when k is finite. In particular, we calculate $\varsigma_1 = p(k-1)(p+2+k)/(4n)$. Applying similar analysis to that in Section A.1.4.1, the conclusion follows by $\vartheta_1(n,p) = \varsigma_1/\sqrt{f}$.

(ii) Chi-squared approximation with the Bartlett correction. In this problem, $\rho = 1-(p+k+2)/(2n)$. It follows that

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)} \left(\frac{2}{\rho n}\right)^l \left[\sum_{j_1=1}^{k-1} B_{l+1}\left\{\frac{(1-\rho)n-(j_1+p)}{2}\right\} - \sum_{j_2=1}^{k-1} B_{l+1}\left\{\frac{(1-\rho)n-j_2}{2}\right\} \right].$$

We calculate that $\varsigma_2 = (k-1)p(p^2+k^2-2k-4)/(48\rho^2n^2)$. Similarly to Section A.1.4.1, the conclusion then follows by $\vartheta_2(n,p) = 2\varsigma_2/\sqrt{f}$.

A.1.4.4 Proof of Theorem 2.2.5 (V): Testing the Equality of Several Covariance Matrices

In this problem, by Section 8.2.4 in Muirhead (2009), we have $f = p(p+1)(k-1)/2$, and

- $K_1 = kp,\ \ K_2 = p;$
- $\xi_{1,j_1} = (n_r - 1)/2, \quad j_1 = (r-1)p + 1, \ldots, rp, \quad (r = 1, \ldots, k);$
- $\upsilon_{1,j_1} = -(r-1)/2, \quad j_1 = r, p + r, \ldots, (k-1)p + r, \ \ (r = 1, \ldots, p);$
- $\xi_{2,j_2} = (n-k)/2,\ \upsilon_{2,j_2} = -(j_2 - 1)/2, \quad j_2 = 1, \ldots, p.$

(i) Chi-squared approximation. Consider $\rho = 1$ and $p^2/n \to 0$. Then

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)}\left[\sum_{r_1=1}^{k}\sum_{r_2=1}^{p}\left(\frac{2}{n_{r_1}-1}\right)^l B_{l+1}\left(-\frac{r_2-1}{2}\right) - \sum_{j=1}^{p}\left(\frac{2}{n-k}\right)^l B_{l+1}\left(-\frac{j-1}{2}\right)\right],$$

which satisfies $\varsigma_l = O(p^{l+2}/n^l)$. Particularly,

$$\varsigma_1 = \left(\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{n-k}\right)\frac{1}{24}p(2p^2+3p-1).$$

Following similar analysis to that in Section A.1.1.2, the conclusion then follows by $\vartheta_1(n,p) = \varsigma_1/\sqrt{f}$.

(ii) Chi-squared approximation with the Bartlett correction. In this problem,

$$\rho = 1 - \frac{(2p^2+3p-1)}{6(p+1)(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{n-k}\right),$$

and we consider $p^3/n^2 \to 0$. In this problem,

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)}\left[\sum_{r_1=1}^{k}\sum_{r_2=1}^{p}\frac{B_{l+1}\{(1-\rho)(n_{r_1}-1)/2-(r_2-1)/2\}}{\{\rho(n_{r_1}-1)/2\}^l}\right.$$
$$\left.-\sum_{j=1}^{p}\frac{B_{l+1}\{(1-\rho)(n-k)/2-(j-1)/2\}}{\{\rho(n-k)/2\}^l}\right].$$

Note that $(1-\rho)(n-k)$ and $(1-\rho)(n_{r_1}-1)$ are of the order of $\Theta(p)$, $B_{l+1}(\cdot)$ is a polynomial of order $l+1$, and k is finite. Then for $l \geq 2$, $\varsigma_l = O(p^{l+2}/n^l)$. In particular, we calculate

$$\varsigma_2 = \frac{p(p+1)}{48\rho^2}\left[(p-1)(p+2)\left\{\sum_{i_1=1}^{k}\frac{1}{(n_{i_1}-1)^2}-\frac{1}{(n-k)^2}\right\}-6(k-1)(1-\rho)^2\right].$$

Similarly to Section A.1.1.2, the conclusion then follows by $\vartheta_2(n,p) = 2\varsigma_2/\sqrt{f}$.

A.1.4.5 Proof of Theorem 2.2.5 (VI): Joint Testing the Equality of Several Mean Vectors and Covariance Matrices

In this problem, by Section 10.8.2 in Muirhead (2009), we have $f = (k-1)p(p+3)/2$, and

- $K_1 = kp$, $K_2 = p$;
- $\xi_{1,j_1} = n_r/2$, $\quad j_1 = (r-1)p+1, \ \ldots, \ rp$, $\quad (r = 1,\ldots,k)$;
- $\upsilon_{1,j_1} = -r/2$, $\quad j_1 = r, \ p+r, \ldots, \ (k-1)p+r$, $\ (r = 1,\ldots,p)$;
- $\xi_{2,j_2} = n/2$, $\upsilon_{2,j_2} = -j_2/2$, $\quad (j_2 = 1,\ldots,p)$.

(i) Chi-squared approximation. Consider $\rho = 1$ and $p^2/n \to 0$. It follows that

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)}\left\{\sum_{r_1=1}^{k}\sum_{r_2=1}^{p}\frac{B_{l+1}(-r_2/2)}{(n_{r_1}/2)^l}-\sum_{j=1}^{p}\frac{B_{l+1}(-j/2)}{(n/2)^l}\right\}.$$

Particularly, we compute

$$\varsigma_1 = \left(\sum_{r=1}^{k} \frac{1}{n_r} - \frac{1}{n} \right) \frac{1}{24} p \left(2p^2 + 9p + 11 \right).$$

Following similar analysis to that in Section A.1.1.2, the conclusion then follows by $\vartheta_1(n,p) = \varsigma_1/\sqrt{f}$.

(ii) Chi-squared approximation with the Bartlett correction. In this problem,

$$\rho = 1 - \left(\sum_{r=1}^{k} \frac{1}{n_r} - \frac{1}{n} \right) \frac{(2p^2 + 9p + 11)}{6(k-1)(p+3)}.$$

It follows that $\varsigma_1 = 0$ and for $l \geq 2$,

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)} \left\{ \sum_{r_1=1}^{k} \sum_{r_2=1}^{p} \frac{B_{l+1}\{(1-\rho)n_{r_1}/2 - r_2/2\}}{(\rho n_{r_1}/2)^l} - \sum_{j=1}^{p} \frac{B_{l+1}\{(1-\rho)n/2 - j/2\}}{(\rho n/2)^l} \right\}.$$

Particularly, we calculate

$$\varsigma_2 = \frac{1}{\rho^2} \left\{ \frac{p(p+1)(p+2)(p+3)}{48} \left(\sum_{i=1}^{k} \frac{1}{n_i^2} - \frac{1}{n^2} \right) - \frac{p(k-1)(p+3)}{8} (1-\rho)^2 \right\}.$$

Applying similar analysis to that in Section A.1.1.2, the conclusion then follows by $\vartheta_2(n,p) = 2\varsigma_2/\sqrt{f}$.

A.1.5 Proofs of Theorems 2.2.3 and 2.2.6

In this section, we prove other problems in Theorems 2.2.3 and 2.2.6 similarly as in Section A.1.1.3. Specifically, we still define $\psi_0(s) = \exp(-s^2/2)$, and we let $\psi_1(s)$ be the characteristic function of $(-2\log\Lambda_n + 2\mu_n)/(2n\sigma_n)$, where Λ_n denotes the corresponding likelihood ratio test statistic, and μ_n and σ_n take the corresponding values given in Theorems 2.2.3 and 2.2.6. By the analysis in Section A.1.1.3, we know that it suffices to prove the results similar to Lemma A.1.4 on Page 149. In particular, in the following subsections, we prove that under H_0 of each test, when $s = o(\min\{(n/p)^{1/2}, f^{1/6}\})$, the characteristic functions satisfy

$$\log \psi_1(s) - \log \psi_0(s) = O\left(\frac{p}{n} + \frac{1}{\sqrt{f}}\right) s + \left(\frac{1}{p} + \frac{p}{n}\right) O\left(s^2\right) + O\left(\frac{s^3}{\sqrt{f}}\right). \quad \text{(A.64)}$$

A.1.5.1 Proof of Theorem 2.2.3 (I): Testing One-Sample Mean Vector

Recall that in Section A.1.2.1, we mention that testing one-sample mean vector can be viewed as testing coefficient vector of a multivariate linear regression model. By Section 10.5.3 in Muirhead (2009), we have

$$\log \psi_1(s) = \log \frac{\Gamma\left\{\frac{1}{2}n(1-ti) - \frac{1}{2}p\right\}}{\Gamma\left\{\frac{1}{2}(n-p)\right\}} - \log \frac{\Gamma\left\{\frac{1}{2}n(1-ti)\right\}}{\Gamma\left(\frac{1}{2}n\right)} + \frac{\mu_n si}{n\sigma_n},$$

where $t = s/(n\sigma_n)$. By (A.7), $t = s/(n\sigma_n) = O(s/\sqrt{f})$. By Lemma A.2.3 (given on Page 184),

$$\begin{aligned} \log \frac{\Gamma\left\{\frac{1}{2}n(1-ti) - \frac{1}{2}p\right\}}{\Gamma\left\{\frac{1}{2}(n-p)\right\}} = & \left\{\frac{1}{2}(n-p) - \frac{1}{2}nti\right\} \log \left\{\frac{1}{2}(n-p) - \frac{1}{2}nti\right\} + \frac{1}{2}nti \\ & - \frac{1}{2}(n-p) \log \left\{\frac{1}{2}(n-p)\right\} + \frac{nti}{2(n-p)} + O\left(\frac{t}{n} + t^2\right). \end{aligned}$$

Similarly, we have

$$\begin{aligned} \log \frac{\Gamma\{\frac{1}{2}n(1-ti)\}}{\Gamma(\frac{1}{2}n)} = & \left\{\frac{n(1-ti)}{2}\right\} \log \left\{\frac{n(1-ti)}{2}\right\} \\ & + \frac{1}{2}nti - \frac{n}{2} \log \left(\frac{n}{2}\right) + \frac{ti}{2} + O\left(\frac{t}{n} + t^2\right). \end{aligned}$$

It follows that

$$\log \psi_1(s) = g_0\left(-\frac{nti}{2}\right) - g_0(0) + \frac{\mu_n si}{n\sigma_n} + O\left(\frac{pt}{n} + t^2\right),$$

where we define in this subsection that $g_0(z) = \{(n-p)/2 + z\} \log\{(n-p)/2 + z\} - (n/2 + z) \log(n/2 + z)$. Following the proof of Lemma A.2.20 (see Section A.2.3.4 on Page 226), we similarly obtain

$$g_0\left(-\frac{nti}{2}\right) - g_0(0) = g_0^{(1)}(0) \times \left(-\frac{nti}{2}\right) - \frac{g_0^{(2)}(0)}{2}\frac{n^2t^2}{4} + O(pt^3),$$

where $g_0^{(1)}(0) = \log\left(1 - \frac{p}{n}\right)$ and $g_0^{(2)}(0) = \frac{2p}{n(n-p)}$. Recall that $2n\sigma_n/\sqrt{2f} \to 1$ by (A.7). Then by Taylor's series and $f = p$,

$$g_0^{(2)}(0)n^2 = 4n^2\sigma_n^2 \left\{1 + O\left(\frac{p}{n}\right)\right\} = 4n^2\sigma_n^2 + O\left(\frac{p^2}{n}\right).$$

Moreover, by Taylor's series, we have $ng_0^{(1)}(0) - 2\mu_n = O(p/n)$. In summary, by $t = s/(n\sigma_n)$ and $n\sigma_n = \Theta(\sqrt{p})$, we obtain

$$\log \psi_1(s) = -\frac{\mu_n si}{n\sigma_n} - \frac{4n^2\sigma_n^2}{2}\frac{s^2}{4(n\sigma_n)^2} + \frac{\mu_n si}{n\sigma_n} + O\left(\frac{ps}{n}\right) + O\left(\frac{p}{n} + \frac{1}{p}\right)s^2 + O\left(\frac{s^3}{\sqrt{p}}\right).$$

Then (A.64) is proved.

A.1.5.2 Proof of Theorem 2.2.3 (II): Testing One-Sample Covariance Matrix

By Corollary 8.3.6 in Muirhead (2009), we have

$$\begin{aligned}\log \psi_1(s) = &-\frac{p(n-1)ti}{2}\log p + \log \frac{\Gamma_p\{\frac{1}{2}(n-1)(1-ti)\}}{\Gamma_p\{\frac{1}{2}(n-1)\}} \\ &+ \log \frac{\Gamma\{\frac{1}{2}p(n-1)\}}{\Gamma\{\frac{1}{2}p(n-1)(1-ti)\}} + \mu_n ti.\end{aligned}$$

By (A.7) and $f = \Theta(p^2)$, $n\sigma_n = \Theta(p)$. Then as $t = s/(n\sigma_n)$, the conditions in Lemma A.2.18 (on Page 221) are satisfied and we have

$$\begin{aligned}\log \frac{\Gamma_p\{(n-1)(1-ti)/2\}}{\Gamma_p\{(n-1)/2\}} = & -\frac{(n-1)\beta_{n,1}ti}{2} + \frac{(n-1)^2\beta_{n,2}t^2}{4} + \beta_{n,3}\Big\{-\frac{(n-1)ti}{2}\Big\} \\ & + O\left(\frac{p^2t}{n}\right) + \left(\frac{1}{p} + \frac{p}{n}\right)O\big(p^2t^2\big) + O\big(p^2t^3\big),\end{aligned}$$

where $\beta_{n,1}$, $\beta_{n,2}$, and $\beta_{n,3}(\cdot)$ are defined in Lemma A.2.18. In addition, we can apply Lemma A.2.3 and obtain

$$\begin{aligned}\log \frac{\Gamma\{p(n-1)/2\}}{\Gamma\{p(n-1)(1-ti)/2\}} = & -p\Big\{\frac{n-1}{2}(1-ti)\Big\}\log\Big[p\Big\{\frac{n-1}{2}(1-ti)\Big\}\Big] \\ & + \frac{p(n-1)}{2}\log\frac{p(n-1)}{2} - \frac{p(n-1)ti}{2} - ti + O\left(\frac{t}{pn} + t^2\right).\end{aligned}$$

By the definition of $\beta_{n,3}(\cdot)$ in Lemma A.2.18, we have

$$\log \frac{\Gamma\{p(n-1)/2\}}{\Gamma\{p(n-1)/2 - pnti/2\}} = -\beta_{n,3}\Big\{-\frac{(n-1)ti}{2}\Big\} - \frac{p(n-1)ti(1-\log p)}{2} + O\left(t + t^2\right).$$

Since $\mu_n = (\beta_{n,1} + p)(n-1)/2$, $2n^2\sigma^2 = \beta_{n,2}(n-1)^2$, $t = s/(n\sigma_n)$, and $n\sigma_n = \Theta(p)$,

$$\log\psi_1(s) - \log\psi_0(s) = O\left(\frac{p}{n} + \frac{1}{p}\right)s + O\left(\frac{1}{p} + \frac{p}{n}\right)s^2 + O\left(\frac{s^3}{p}\right).$$

A.1.5.3 Proof of Theorem 2.2.6 (IV): Testing the Equality of Several Mean Vectors

By (A.57) and the analysis in Section A.1.4.3, we have

$$\log\psi_1(s) = \sum_{j=1}^{k-1}\left[\log\frac{\Gamma\big\{\frac{1}{2}(n-j-p) - \frac{1}{2}nti\big\}}{\Gamma\big\{\frac{1}{2}(n-j-p)\big\}} - \log\frac{\Gamma\big\{\frac{1}{2}(n-j) - \frac{1}{2}nti\big\}}{\Gamma\big\{\frac{1}{2}(n-j)\big\}}\right] + \frac{\mu_n si}{n\sigma_n},$$

where $t = s/(n\sigma_n)$. By Lemma A.2.3,

$$\log \frac{\Gamma\{\frac{1}{2}(n-j-p) - \frac{1}{2}nti\}}{\Gamma\{\frac{1}{2}(n-j-p)\}} = \left\{\frac{1}{2}(n-j-p) - \frac{1}{2}nti\right\} \log\left\{\frac{1}{2}(n-j-p) - \frac{1}{2}nti\right\} \\ - \frac{n-j-p}{2} \log \frac{n-j-p}{2} + \frac{nti}{2} + O(t+t^2).$$

Applying similar analysis, we obtain

$$\log \frac{\Gamma\{\frac{1}{2}(n-j-nti)\}}{\Gamma\{\frac{1}{2}(n-j)\}} = \left(\frac{n-j-nti}{2}\right) \log\left(\frac{n-j-nti}{2}\right) \\ - \frac{n-j}{2} \log \frac{n-j}{2} + \frac{nti}{2} + O(t+t^2).$$

It follows that $\log \psi_1(s) = \sum_{j=1}^{k-1} \{g_j(nti/2) - g_j(0)\} + \mu_n si/(n\sigma_n) + O(t+t^2)$, where we define in this subsection that

$$g_j(z) = \left(\frac{n-j-p}{2} - z\right) \log\left(\frac{n-j-p}{2} - z\right) - \left(\frac{n-j}{2} - z\right) \log\left(\frac{n-j}{2} - z\right).$$

Following similar proof to that of Lemma A.2.20 (see Section A.2.3.4), we obtain

$$\sum_{j=1}^{k-1} \{g_j(nti) - g_j(0)\} = \sum_{j=1}^{k-1} g_j^{(1)}(0) \frac{nti}{2} - \frac{n^2t^2}{8} \sum_{j=1}^{k-1} g_j^{(2)}(0) + O(pt^3), \tag{A.65}$$

where $g_j^{(1)}(0) = \log(\frac{n-j}{2}) - \log(\frac{n-j-p}{2})$ and $g_j^{(2)}(0) = \frac{2}{n-j-p} - \frac{2}{n-j}$. Note that

$$\frac{1}{2} \sum_{j=1}^{k-1} g_j^{(2)}(0) = \sum_{j=1}^{k-1} \frac{p}{(n-j-p)(n-j)} = \frac{p(k-1)}{(n-p-1)n} \left\{1 + O\left(\frac{k}{n}\right)\right\},$$

and

$$2\sigma_n^2 = \log\left\{1 + \frac{p(k-1)}{(n-k)(n-p-1)}\right\} = \frac{p(k-1)}{(n-p-1)n} \left\{1 + O\left(\frac{k}{n}\right)\right\}.$$

Thus $\sum_{j=1}^{k-1} g_j^{(2)}(0)(4\sigma_n^2)^{-1} = 1 + O(n^{-1})$. In addition,

$$\sum_{j=1}^{k-1} g_j^{(1)}(0) = \log \frac{\Gamma(n-1)}{\Gamma(n-k)} - \log \frac{\Gamma(n-p-1)}{\Gamma(n-p-k)}.$$

We then apply Lemma A.2.1 to expand the $\log \Gamma(\cdot)$ function, and calculate

$$\begin{aligned}\sum_{j=1}^{k-1} g_j^{(1)}(0) = & -\left(n-p-k-\frac{1}{2}\right)\left\{\log\left(1-\frac{p}{n-1}\right) - \log\left(1-\frac{p}{n-k}\right)\right\} \\ & - p\log\left(1-\frac{k-1}{n-1}\right) - (k-1)\log\left(1-\frac{p}{n-1}\right) + O(n^{-1}).\end{aligned}$$

Therefore $\sum_{j=1}^{k-1} g_j^{(1)}(0) = -\mu_n/n + O(n^{-1})$. Then by (A.65), $t = s/(n\sigma_n)$, $n\sigma_n = \Theta(f^{1/2})$, and $f = \Theta(p)$, we have

$$\begin{aligned}\log \psi_1(s) &= \big\{ -\mu_n/n + O(n^{-1})\big\} nti - \frac{n^2\sigma_n^2 t^2}{2}\big\{1 + O(n^{-1})\big\} + \mu_n ti + O\left(t + t^2 + pt^3\right) \\ &= -\frac{s^2}{2} + O\left(\frac{1}{\sqrt{f}}\right) s + O\left(\frac{p}{n} + \frac{1}{f}\right) s^2 + O\left(\frac{s^3}{\sqrt{f}}\right).\end{aligned}$$

By $\log \psi_0(s) = -s^2/2$, (A.64) is proved.

A.1.5.4 Proof of Theorem 2.2.6 (V): Testing the Equality of Several Covariance Matrices

By (A.57) and the analysis in Section A.1.4.4, we have

$$\begin{aligned}\log \psi_1(s) = & \log \frac{\Gamma_p\big\{\frac{1}{2}(n-k)\big\}}{\Gamma_p\big\{\frac{1}{2}(n-k)(1-ti)\big\}} + \sum_{j=1}^{k} \log \frac{\Gamma_p\big\{\frac{1}{2}(n_j-1)(1-ti)\big\}}{\Gamma_p\big\{\frac{1}{2}(n_j-1)\big\}} \\ & - p\left\{(n-k)\log(n-k) - \sum_{j=1}^{k}(n_j-1)\log(n_j-1)\right\}\frac{ti}{2} + \frac{\mu_n si}{n\sigma_n},\end{aligned}$$

where $t = s/(n\sigma_n)$. By Lemma A.2.18, we can expand $\log \Gamma_p(\cdot)$ and obtain

$$\log \psi_1(s) = -\mu_n ti - \frac{n^2\sigma_n^2 t^2}{2} + \mu_n ti + R_n(t), \tag{A.66}$$

where the calculations of μ_n and σ_n are similar to that in Section A.5 of Jiang and Qi (2015), and thus the details are skipped here. In (A.66), $R_n(t)$ denotes the remainder term of the expansion. Since Lemma A.2.18 is used, we know that the remainder term satisfies

$$R_n(t) = O\left(\frac{p}{n}\right) s + \left(\frac{1}{p} + \frac{p}{n}\right) s^2 + O\left(\frac{s^3}{p}\right).$$

By $t = s/(n\sigma_n)$ and (A.66), (A.64) is obtained.

A.1.5.5 Proof of Theorem 2.2.6 (VI): Joint Testing the Equality of Several Mean Vectors and Covariance Matrices

By Corollary 10.8.3 in Muirhead (2009),

$$\begin{aligned}\log \psi_1(s) = {} & \log \frac{\Gamma_p\{\frac{1}{2}(n-1)\}}{\Gamma_p\{\frac{1}{2}(n-1) - \frac{1}{2}nti\}} + \sum_{j=1}^{k} \log \frac{\Gamma_p\{\frac{1}{2}(n_j-1) - \frac{1}{2}n_j ti\}}{\Gamma_p\{\frac{1}{2}(n_j-1)\}} \\ & - p\left(n\log n - \sum_{j=1}^{k} n_j \log n_j\right)\frac{ti}{2} + \frac{\mu_n si}{n\sigma_n},\end{aligned}$$

where $t = s/(n\sigma_n)$. By Lemma A.2.18,

$$\begin{aligned} & \log \frac{\Gamma_p\{\frac{1}{2}(n_j-1) - \frac{1}{2}n_j ti\}}{\Gamma_p\{\frac{1}{2}(n_j-1)\}} \\ = {} & \left[2pn_j + \left(n_j - p - \frac{3}{2}\right) n_j \log\left(1 - \frac{p}{n_j-1}\right)\right]\frac{ti}{2} \\ & + \left\{\frac{p}{n_j-1} + \log\left(1 - \frac{p}{n_j-1}\right)\right\}\frac{n_j^2 t^2}{4} + \varrho_{n_j}(t) + R_n(t), \end{aligned} \tag{A.67}$$

where for an integer l, we define

$$\varrho_l(t) = p\left\{\left(\frac{l-1}{2}+\frac{lt}{2}\right)\log\left(\frac{l-1}{2}+\frac{lt}{2}\right)-\frac{l-1}{2}\log\frac{l-1}{2}\right\}, \tag{A.68}$$

and $R_n(t)$ denotes the remainder term and it is of the order of

$$R_n(t) = O\left(\frac{pt}{n}\right)+O\left(\frac{1}{p}+\frac{p}{n}\right)p^2t^2+O\big(p^2t^3\big). \tag{A.69}$$

In addition, to evaluate $\log\psi_1(s)$, we also use Lemma A.1.6 below.

Lemma A.1.6. *Under the conditions of Theorem 2.2.6, as $p/n \to 0$ and $t = s/(n\sigma_n) = O(s/\sqrt{f})$,*

$$\begin{aligned} n^2t^2\log\left(1-\frac{p}{n-1}\right) =& n^2t^2\log\left(1-\frac{p}{n}\right)+O\left(\frac{p}{n}\right)t^2, \qquad \text{(A.70)}\\ \left\{(n-p-\frac{3}{2})n\log\left(1-\frac{p}{n-1}\right)\right\}t =& \left\{(n-p-\frac{3}{2})n\log\left(1-\frac{p}{n}\right)\right\}t-pt+O\left(\frac{pt}{n}\right). \end{aligned}$$

Moreover, for $\varrho_l(t)$ defined in (A.68), *we have*

$$-\varrho_n(t)+\sum_{j=1}^{k}\varrho_{n_j}(t) = \left(1-k-n\log n+\sum_{j=1}^{k}n_j\log n_j\right)\frac{tp}{2}+O\left(\frac{pt}{n}+pt^2\right). \tag{A.71}$$

Proof. Please see Section A.2.3.5 on Page 227. $\square$

By Lemma A.1.6 and the expansions of gamma functions in (A.67), we calculate

$$
\begin{aligned}
&\log \psi_1(s) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{(A.72)}\\
&= \left\{p - \left(n - p - \frac{3}{2}\right) n \log\left(1 - \frac{p}{n}\right) + \sum_{j=1}^{k} \left(n_j - p - \frac{3}{2}\right) n_j \log\left(1 - \frac{p}{n_j - 1}\right)\right\} \frac{ti}{2}\\
&\quad - \left(n^2 L_{n,p} - \sum_{j=1}^{k} n_j^2 L_{n_j-1,p}\right) \frac{t^2}{4} - p\left\{(1-k) - n\log n + \sum_{j=1}^{k} n_j \log n_j\right\} \frac{ti}{2}\\
&\quad - p\left(n \log n - \sum_{j=1}^{k} n_j \log n_j\right) \frac{ti}{2} + \frac{\mu_n s i}{n\sigma_n} + R_n(t),
\end{aligned}
$$

where $R_n(t)$ denotes the remainder term of (A.72), which is of the order same as that in (A.69), whereas we mention that the exact value of $R_n(t)$ can change. Then we obtain (A.64) by $t = s/(n\sigma_n)$ and $n\sigma_n = \Theta(f^{1/2})$.

A.2 Proofs of Technical Lemmas in Section A.1

A.2.1 Asymptotic Expansions of the Gamma Functions

In this section, we provide some results on asymptotic expansions of the gamma functions, which are repeatedly used in the proofs. We first give the following Lemma A.2.1 on the expansion of $\log \Gamma(z)$, which also provides the basis for other lemmas below. Lemma A.2.1 and its proof can be found in 12.33 of Whittaker and Watson (1996).

Lemma A.2.1. *Suppose that a complex number z satisfies $Re(z) \geq \epsilon_1 > 0$ and $|\arg(z)| \leq \pi/2 - \epsilon_2$ with $\epsilon_1 > 0$ and $0 < \epsilon_2 < \pi/4$ being given in advance. When $|z| \to \infty$, and an even integer L, we have*

$$
\log \Gamma(z) = \left(z - \frac{1}{2}\right) \log z - z + \log \sqrt{2\pi} + \sum_{l=1}^{L-1} \frac{(-1)^{l+1} B_{l+1}(0)}{l(l+1)z^l} + R_L(z), \qquad \text{(A.73)}
$$

where $B_{l+1}(\cdot)$ represents the Bernoulli polynomial of order $l+1$, and

$$|R_L(z)| = O\left(\frac{|B_{L+2}(0)|}{(L+1)(L+2)|z|^{L+1}}\right).$$

Particularly, we know $B_l(0) = 0$ when l is odd and $l \geq 3$.

In Lemma A.2.1, if we take $L = 2$ and z as a real number, by $B_2(0) = 1/6$, we have

$$\log \Gamma(z) = \left(z - \frac{1}{2}\right)\log z - z + \log\sqrt{2\pi} + \frac{1}{12z} + O(z^{-2}). \tag{A.74}$$

Given Lemma A.2.1, we next prove two additional lemmas on asymptotic expansions of the gamma functions.

Lemma A.2.2. *Suppose a complex number $z+a$ satisfies $Re(z+a) \geq \epsilon_1 > 0$ and $|\arg(z+a)| \leq \pi/2 - \epsilon_2$ with $\epsilon_1 > 0$ and $0 < \epsilon_2 \leq \pi/4$ being given in advance. Assume $|a| \to \infty$ as $|z| \to \infty$ and $|a| = o(|z|)$. For a finite even L, when $|a|^{L+1}/|z|^L \to 0$,*

$$\log\Gamma(z+a) = \left(z + a - \frac{1}{2}\right)\log z - z + \log\sqrt{2\pi} + \sum_{l=1}^{L-1}\frac{(-1)^{l+1}B_{l+1}(a)}{l(l+1)z^l} + O\left(\frac{|a|^{L+1}}{|z|^L}\right).$$

Proof. Please see Section A.2.1.1 on Page 185. □

Lemma A.2.3. *For a real number $x \to \infty$ and a real number $b = o(x)$,*

$$\log\frac{\Gamma(x+bi)}{\Gamma(x)} = (x+bi)\log(x+bi) - x\log x - bi - \frac{bi}{2x} + O\left(\frac{b+b^2}{x^2}\right),$$

where i denotes the imaginary unit.

Proof. Please see Section A.2.1.2 on Page 186. □

A.2.1.1 Proof of Lemma A.2.2 (on Page 184)

By (A.73), for a finite even L, we have

$$
\begin{aligned}
&\log\Gamma(z+a) \qquad\qquad (A.75)\\
=\ & \left(z+a-\frac{1}{2}\right)\log(z+a)-(z+a)+\log\sqrt{2\pi}+\sum_{l=1}^{L-1}\frac{(-1)^{l+1}B_{l+1}(0)}{l(l+1)(z+a)^l}+O\big(|z+a|^{-L-1}\big)\\
=\ & \left(z+a-\frac{1}{2}\right)\log z - z+\left(z+a-\frac{1}{2}\right)\log\left(1+\frac{a}{z}\right)-a+\log\sqrt{2\pi}\\
& +\sum_{l=1}^{L-1}\frac{(-1)^{l+1}B_{l+1}(0)}{l(l+1)z^l}\left(1+\frac{a}{z}\right)^{-l}+O\big(|z+a|^{-L-1}\big).
\end{aligned}
$$

By Taylor's expansion,

$$
\left(z+a-\frac{1}{2}\right)\log\left(1+\frac{a}{z}\right)-a=\sum_{k=1}^{L-1}\frac{(-1)^{k+1}}{z^k}\left\{\frac{a^{k+1}}{k(k+1)}-\frac{1}{2k}a^k\right\}+O\left(\frac{|a|^{L+1}}{|z|^L}\right). \tag{A.76}
$$

Note that $B_0(0)=1$ and $B_1(0)=-1/2$. Thus

$$
(A.76)=\sum_{k=1}^{L-1}\frac{(-1)^{k+1}}{k(k+1)z^k}\left\{B_0(0)a^{k+1}+\binom{k+1}{1}B_1(0)a^k\right\}+O\left(\frac{|a|^{L+1}}{|z|^L}\right). \tag{A.77}
$$

In addition, by Taylor's expansion, when L is finite,

$$
\begin{aligned}
&\sum_{l=1}^{L-1}\frac{(-1)^{l+1}B_{l+1}(0)}{l(l+1)z^l}\left(1+\frac{a}{z}\right)^{-l} \qquad\qquad (A.78)\\
=\ & \sum_{l=1}^{L-1}\frac{(-1)^{l+1}B_{l+1}(0)}{l(l+1)z^l}\left\{\sum_{s=0}^{L-1-l}(-1)^s\binom{l+s-1}{s}\frac{a^s}{z^s}+O\left(|a/z|^{L-l}\right)\right\}\\
=\ & \sum_{k=1}^{L-1}\sum_{t=1}^{k}\frac{(-1)^{k+1}B_{t+1}(0)}{t(t+1)z^k}\frac{(k-1)!}{(t-1)!(k-t)!}a^{k-t}+O\left(|a/z|^L\right)\\
=\ & \sum_{k=1}^{L-1}\sum_{t=2}^{k+1}\frac{(-1)^{k+1}B_t(0)}{k(k+1)z^k}\binom{k+1}{t}a^{k+1-t}+O\left(|a/z|^L\right).
\end{aligned}
$$

Combining (A.75), (A.77), and (A.78), we obtain

$$\log\Gamma(z+a) = \left(z+a-\frac{1}{2}\right)\log z - z + \log\sqrt{2\pi}$$
$$+\sum_{k=1}^{L-1}\frac{(-1)^{k+1}}{k(k+1)z^k}\left\{\sum_{t=0}^{k+1}\binom{k+1}{t}B_t(0)a^{k+1-t}\right\} + O\left(\frac{|a|^{L+1}}{|z|^L}\right).$$

By the property of the Bernoulli polynomials, $B_{k+1}(a) = \sum_{t=0}^{k+1} B_t(0)a^{k+1-t}$; see, e.g., Eq. (13) on Page 21 in Luke (1969). Therefore the lemma is proved.

A.2.1.2 Proof of Lemma A.2.3 (on Page 184)

By Binet's second formula of the gamma function, it can be obtained that for a complex number z with positive real part, and any integer $L \geq 1$,

$$\log\Gamma(z) = \left(z-\frac{1}{2}\right)\log z - z + \log\sqrt{2\pi} + \sum_{l=1}^{L}\frac{B_{2l}(0)}{(2l-1)(2l)z^{2l-1}}$$
$$+\frac{2(-1)^L}{z^{2L-1}}\int_0^\infty\int_0^t\frac{u^{2L}\mathrm{d}u}{u^2+z^2}\frac{\mathrm{d}t}{e^{2\pi t}-1};$$

please see Page 252 in Whittaker and Watson (1996) for details. Take $L=1$, and by $B_2(0) = 1/6$, we have

$$\log\Gamma(x) = \left(x-\frac{1}{2}\right)\log x - x + \log\sqrt{2\pi} + \frac{1}{12x} - \frac{2}{x}\int_0^\infty\left(\int_0^t\frac{u^2}{x^2+u^2}\mathrm{d}u\right)\frac{\mathrm{d}t}{e^{2\pi t}-1}.$$

Similarly, we have

$$\log\Gamma(x+bi) = \left(x+bi-\frac{1}{2}\right)\log(x+bi) - (x+bi) + \log\sqrt{2\pi}$$
$$+\frac{1}{12(x+bi)} - \frac{2}{x+bi}\int_0^\infty\left(\int_0^t\frac{u^2}{(x+bi)^2+u^2}\mathrm{d}u\right)\frac{\mathrm{d}t}{e^{2\pi t}-1}.$$

It follows that

$$\log \frac{\Gamma(x+bi)}{\Gamma(x)} \tag{A.79}$$
$$= (x+bi)\log(x+bi) - x\log x - bi - \frac{1}{2}\log\left(1+\frac{bi}{x}\right) + \frac{1}{12}\left(\frac{1}{x+bi} - \frac{1}{x}\right) + \tilde{R}_2,$$

where

$$\tilde{R}_2 = -2\int_0^\infty \int_0^t \left[\frac{u^2}{(x+bi)\{(x+bi)^2+u^2\}} - \frac{u^2}{x(x^2+u^2)}\right] \mathrm{d}u \frac{\mathrm{d}t}{e^{2\pi t}-1}.$$

To evaluate $\tilde{R}_2$, we note that

$$\begin{aligned} &\frac{u^2}{(x+bi)\{(x+bi)^2+u^2\}} - \frac{u^2}{x(x^2+u^2)} \\ = &-\frac{u^2 b_x}{x^3(1+b_x i)(1+u_x^2)} \times \left[\frac{2i-b_x}{(1+b_x i)^2+u_x^2} + i\right], \end{aligned}$$

where for easy presentation, we let $b_x = b/x$ and $u_x = u/x$. Since $b = o(x)$, $|(1+b_x i)^{-1}|$ is bounded. Moreover, we also know $(1+u_x^2)^{-1}$ and $|\{(1+b_x i)^2+u_x^2\}^{-1}|$ are bounded. It follows that there exists a constant C such that

$$|\tilde{R}_2| \leq \frac{Cb_x}{x^3}\int_0^\infty \left(\int_0^t u^2 \mathrm{d}u\right) \frac{\mathrm{d}t}{e^{2\pi t}-1} = O\left(\frac{b}{x^4}\right),$$

where we use $\int_0^\infty t^3(e^{2\pi t}-1)^{-1}\mathrm{d}t$ is a constant; see 7.2 in Whittaker and Watson (1996). Lemma A.2.3 is then obtained by (A.79) and $\log(1+bi/x) = bi/x + O(b^2/x^2)$ and $(x+bi)^{-1} - x^{-1} = O(bx^{-2})$.

A.2.2 Lemmas for Theorems 2.2.2 and 2.2.5

A.2.2.1 Proof of Lemma A.1.1 (on Page 143)

By (A.1), we can write $\log \mathrm{E}\{\exp(-2it\eta \log \Lambda_n)\} = G_1 + G_2 + G_3$, where in this subsection, we let

$$
\begin{aligned}
&G_1 = -i\eta npt \log\left(\frac{2e}{n}\right), \quad G_2 = -\frac{np}{2}(1-2i\eta t)\log(1-2i\eta t),\\
&G_3 = \log \Gamma_p\left(\frac{n-1}{2} - \eta nit\right) - \log \Gamma_p\Big(\frac{n-1}{2}\Big).
\end{aligned}
$$

By the property of multivariate gamma function; see, e.g., Theorem 2.1.12 in Muirhead (2009), we obtain

$$
\begin{aligned}
G_3 &= \sum_{j=1}^{p} \log \Gamma\left\{\frac{n}{2}(1-2\eta it) - \frac{j}{2}\right\} - \sum_{j=1}^{p} \log \Gamma\left(\frac{n}{2} - \frac{j}{2}\right)\\
&= \sum_{j=1}^{p} \left[\log \Gamma\left\{\frac{\eta n}{2}(1-2it) + \frac{n(1-\eta)-j}{2}\right\} - \log \Gamma\left\{\frac{\eta n}{2} + \frac{n(1-\eta)-j}{2}\right\}\right].
\end{aligned}
$$

We first examine G_3. When $\eta = 1$ or $\eta = \rho$, for $1 \le j \le p$, $n(1-\eta) - j = O(p)$ and $\eta n = \Theta(n)$. As $p = o(n)$, $|\{n(1-\eta)-j\}\{\eta n(1-2it)\}^{-1}| = O(p/n) = o(1)$. Then we can apply Lemma A.2.2 on Page 184, and obtain

$$
\begin{aligned}
&\log \Gamma\left\{\frac{\eta n}{2}(1-2it) + \frac{n(1-\eta)-j}{2}\right\}\\
=& \left\{\frac{\eta n}{2}(1-2it) + \frac{n(1-\eta)-j-1}{2}\right\} \log\left\{\frac{\eta n}{2}(1-2it)\right\} - \frac{\eta n}{2}(1-2it) + \log\sqrt{2\pi}\\
&+ \sum_{l=1}^{L-1} \frac{(-1)^{l+1}}{l(l+1)} B_{l+1}\left\{\frac{n(1-\eta)}{2} - \frac{j}{2}\right\}\left\{\frac{\eta n}{2}(1-2it)\right\}^{-l} + O\Big(\frac{p^{L+1}}{n^L}\Big),
\end{aligned}
$$

and

$$
\begin{aligned}
& \log \Gamma\left\{\frac{\eta n}{2}+\frac{n(1-\eta)-j}{2}\right\} \\
= & \left\{\frac{\eta n}{2}+\frac{n(1-\eta)-j-1}{2}\right\} \log \frac{\eta n}{2}-\frac{\eta n}{2}+\log \sqrt{2 \pi} \\
& +\sum_{l=1}^{L-1} \frac{(-1)^{l+1}}{l(l+1)} B_{l+1}\left\{\frac{n(1-\eta)}{2}-\frac{j}{2}\right\}\left(\frac{\eta n}{2}\right)^{-l}+O\left(\frac{p^{L+1}}{n^{L}}\right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
G_3= & -\eta p n t i \log \left(\frac{n}{2 e}\right)-p \eta n i t \log \eta+\frac{p n}{2}(1-2 i \eta t) \log (1-2 i t)-\sum_{j=1}^{p} \frac{j+1}{2} \log (1-2 i t) \\
& +\sum_{l=1}^{L-1} \frac{(-1)^{l+1}}{l(l+1)} \sum_{j=1}^{p} B_{l+1}\left\{\frac{n(1-\eta)}{2}-\frac{j}{2}\right\}\left(\frac{\eta n}{2}\right)^{-l}\left\{(1-2 i t)^{-l}-1\right\}+O\left(\frac{p^{L+2}}{n^{L}}\right).
\end{aligned}
$$

We next examine G_2. By $1-2i\eta t=\eta(1-2it)+1-\eta$, and Taylor's expansion,

$$
\begin{aligned}
& (1-2 i \eta t) \log (1-2 i \eta t) \\
= & \{\eta(1-2 i t)+1-\eta\} \log \{\eta(1-2 i t)\} \\
& +1-\eta+(1-\eta) \sum_{l=1}^{L-1} \frac{(-1)^{l+1}}{l(l+1)}\left(\frac{1-\eta}{\eta}\right)^{l}(1-2 i t)^{-l}+O\left\{(1-\eta)^{L+1}\right\}.
\end{aligned}
$$

As $\log(1)=(1-2i\eta\times 0)\log(1-2i\eta\times 0)=0$, by applying Taylor's expansion similarly as above,

$$
\begin{aligned}
& (1-2 i \eta t) \log (1-2 i \eta t)-\log (1) \\
= & -2 i \eta t \log \eta(1-2 i t)+\log (1-2 i t) \\
& +(1-\eta) \sum_{l=1}^{L-1} \frac{(-1)^{l+1}}{l(l+1)}\left(\frac{1-\eta}{\eta}\right)^{l}\left\{(1-2 i t)^{-l}-1\right\}+O\left\{(1-\eta)^{L+1}\right\}.
\end{aligned}
$$

As $(1-\eta)/\eta = \{(1-\eta)n/2\}/(\eta n/2)$,

$$G_2 = -\sum_{l=1}^{L-1} \frac{(-1)^{l+1}}{l(l+1)} \sum_{j=1}^{p} \left\{ \frac{(1-\eta)n}{2} \right\}^{l+1} \left(\frac{\eta n}{2} \right)^{-l} \left\{ (1-2it)^{-l} - 1 \right\}$$
$$+ i\eta npt \log \eta (1-2it) - \frac{np}{2} \log(1-2it) + O\left\{ (1-\eta)^{L+1} pn \right\}.$$

In summary, as $1-\eta = O(p/n)$ when $\eta = 1$ or ρ, we have

$$G_1 + G_2 + G_3 = -\sum_{j=1}^{p} \frac{j+1}{2} \log(1-2it) + \sum_{l=1}^{L-1} \varsigma_l \left\{ (1-2it)^{-l} - 1 \right\} + O\left(\frac{p^{L+2}}{n^L} \right),$$

where

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)} \sum_{j=1}^{p} \left[B_{l+1} \left\{ \frac{(1-\eta)n}{2} - \frac{j}{2} \right\} - \left\{ \frac{(1-\eta)n}{2} \right\}^{l+1} \right] \left(\frac{\eta n}{2} \right)^{-l}.$$

Particularly, as $B_{l+1}(\cdot)$ is a polynomial of order $l+1$ and $(1-\eta)n = O(p)$, we have $\varsigma_l = O(p^{l+2} n^{-l})$.

A.2.2.2 Notation of the finite difference and computation rules

In the following, we prove Propositions A.1.1 and A.1.2 and Lemma A.1.2 based on the calculus of the finite difference. To facilitate the proofs, we introduce some notation. Given x, define a function with respect to the degrees of freedom f as $F_x(f) = P(\chi^2_f \leq x)$. Let Δ_{2h} represent a forward difference operator with step $2h$, that is, $\Delta_{2h}(F_x, f) = F_x(f+2h) - F_x(f)$. For an integer $v \geq 1$, it follows that the v-th order forward difference is $\Delta_{2h}^v(F_x, f) = \sum_{w=0}^{v} \binom{v}{w} (-1)^{v-w} F(f + 2hw)$, where $\Delta_{2h}^1(F_x, f) = \Delta_{2h}(F_x, f)$. Particularly, when $h = 1$, we have $\Delta_2^v(F_x, f) = \sum_{w=0}^{v} \binom{v}{w} (-1)^{v-w} P(\chi^2_{f+2w} \leq x)$; when $h = 2$, $\Delta_4^v(F_x, f) = \sum_{w=0}^{v} \binom{v}{w} (-1)^{v-w} P(\chi^2_{f+4w} \leq x)$. In the following proofs, we use several rules of the finite difference operator listed in Lemmas A.2.4–A.2.6 below, which can be found in Section 3.7 of Zwillinger (2002).

Lemma A.2.4 (Leibniz rule)**.** *For two functions $F(f)$ and $G(f)$, and two positive integers v and h, $\Delta_h^v(FG, f) = \sum_{w=0}^{v} \binom{v}{w} \Delta_h^w(F, f)\Delta_h^{v-w}(G, f + hw)$.*

Lemma A.2.5 (Linearity rule)**.** *For two constants C_1 and C_2, two functions $F(f)$ and $G(f)$, and two positive integers v and h, the linear combination $C_1F(f)+C_2G(f)$ satisfies $\Delta_h^v(C_1F + C_2G, f) = C_1\Delta_h^v(F) + C_2\Delta_h^v(G)$.*

Lemma A.2.6. *For a function $F(f)$ and positive integers v_1, v_2, h_1, and h_2,*

$$\Delta_{h_2}^{v_2}\Delta_{h_1}^{v_1}(F, f) = \Delta_{h_1}^{v_1}\Delta_{h_2}^{v_2}(F, f) = \Delta_{h_2}^{v_2}\Delta_{h_1}^{v_1-1}\{\Delta_{h_1}(F, f)\} = \Delta_{h_1}^{v_1}\Delta_{h_2}^{v_2-1}\{\Delta_{h_2}(F, f)\}.$$

Based on the notation and lemmas on the finite difference, we first prove Lemma A.1.2 in Section A.2.2.3, and then use Lemma A.1.2 to prove Propositions A.1.1 and A.1.2 in Sections A.2.2.4 and A.2.2.5, respectively.

A.2.2.3 Proof of Lemma A.1.2 (on Page 146)

We prove (A.24) in Lemma A.1.2 from the cumulative distribution function of the chi-squared distribution. In particular, by the probability density of χ_f^2, we have $\Pr\left(\chi_f^2 \le x\right) = \gamma(f/2, x/2)/\Gamma(f/2)$, where $\gamma(m, x)$ is the lower incomplete gamma function defined as $\gamma(m, x) = \int_0^x t^{m-1}e^{-t}\mathrm{d}t$, and $\Gamma(m)$ is the gamma function defined as $\Gamma(m) = \int_0^\infty t^{m-1}e^{-t}\mathrm{d}t$; see, e.g., Section 6.2 in Press et al. (1992). Thus for an integer h,

$$\Delta_{2h}^1(F_x, f) = \frac{\Gamma\left(\frac{f}{2}\right)\gamma\left(\frac{f}{2} + h, \frac{x}{2}\right) - \Gamma\left(\frac{f}{2} + h\right)\gamma\left(\frac{f}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{f}{2} + h\right)\Gamma\left(\frac{f}{2}\right)},$$

where $\Delta_{2h}^1(F_x, f) = \Pr(\chi_{f+2h}^2 \le x) - \Pr(\chi_f^2 \le x)$ following the notation in Section A.2.2.2. By integration by parts, we have

$$\Gamma(m + 1) = m\Gamma(m), \qquad \text{and then} \qquad \Gamma(m + h) = \prod_{k=1}^{h}(m + h - k)\Gamma(m). \tag{A.80}$$

Similarly, we have $\gamma(m+1,x) = m\gamma(m,x) - x^m e^{-x}$, and then

$$\gamma(m+h,x) = \prod_{k=1}^{h}(m+h-k)\gamma(m,x) - \sum_{k=1}^{h}\prod_{t=1}^{k-1}(m+h-t)x^{m+h-k}e^{-x};$$

this recurrence formulas can also be found in Sections 6.3 and 6.5 in Abramowitz and Stegun (1970). It follows that

$$\Delta_{2h}^{1}(F_x,f) = -\frac{\sum_{k=1}^{h}\prod_{t=1}^{k-1}(f/2+h-t)(x/2)^{\frac{f}{2}+h-k}e^{-x/2}}{\prod_{t=1}^{h}(f/2+h-t)\times\Gamma(f/2)} = -\sum_{k=1}^{h}\frac{(x/2)^{\frac{f}{2}+h-k}\,e^{-x/2}}{\Gamma\left(f/2+h-k+1\right)}.$$

Therefore (A.24) is proved.

We next prove (A.25) in Lemma A.1.2 based on (A.24) by discussing $h \in \{1,2,3,4\}$, respectively.

(1). We first consider $h=1$. Under this case,

$$\Delta_{2}^{1}(F_x,f) = -\frac{(x/2)^{f/2}e^{-x/2}}{\Gamma(f/2+1)}. \tag{A.81}$$

By (A.74), as $f\to\infty$, $\Gamma(f/2) = (f/2)^{f/2-1/2}e^{-f/2}\sqrt{2\pi}\{1+O(f^{-1})\}$. Moreover, by $\Gamma(f/2+1) = \Gamma(f/2)f/2$, we have

$$\begin{aligned}\frac{1}{\Gamma(f/2+1)}(x/2)^{f/2}e^{-x/2} &= \frac{1}{\sqrt{f\pi}}\left(\frac{x}{f}\right)^{f/2}\exp\left\{\frac{f-x}{2}+O(f^{-1})\right\}\\ &= \frac{1}{\sqrt{f\pi}}\exp\left\{\frac{f-x}{2}+\frac{f}{2}\log\left(1+\frac{x-f}{f}\right)+O(f^{-1})\right\}.\end{aligned}$$

When $x=\chi_f^2(\alpha)$, we have $x = f+\sqrt{2f}\{z_\alpha+O(f^{-1/2})\}$ by (A.6). Then by Taylor's series,

$$\Delta_{2}^{1}(F_x,f) = \frac{1}{\sqrt{f\pi}}\exp\left\{-\frac{(x-f)^2}{4f}+O(f^{-1/2})\right\} = \frac{1}{\sqrt{f\pi}}\exp\left(-\frac{z_\alpha^2}{2}\right)\{1+O(f^{-1/2})\}.$$

(2). When $h = 2$, by (A.24), (A.80), and $x = f + \sqrt{2f}\{z_\alpha + O(f^{-1/2})\}$, we have

$$\Delta_4^1(F_x, f) = -\frac{\frac{x}{2}}{\frac{f}{2}+1} \times \Delta_2^1(F_x, f) + \Delta_2^1(F_x, f) = -\frac{2}{\sqrt{f\pi}} \exp\left(-\frac{z_\alpha^2}{2}\right)\{1 + O(f^{-1/2})\}.$$

(3). When $h = 3$, similarly by (A.24), (A.80), and $x = f + \sqrt{2f}\{z_\alpha + O(f^{-1/2})\}$, we have

$$\Delta_6^1(F_x, f) = -\frac{(\frac{x}{2})^2}{(\frac{f}{2}+2)(\frac{f}{2}+1)}\Delta_2^1(F_x, f) + \Delta_4^1(F_x, f) = -\frac{3}{\sqrt{f\pi}} \exp\left(-\frac{z_\alpha^2}{2}\right)\{1 + O(f^{-1/2})\}.$$

(4). When $h = 4$, similarly by (A.24), (A.80), and $x = f + \sqrt{2f}\{z_\alpha + O(f^{-1/2})\}$, we have

$$\begin{aligned}\Delta_8^1(F_x, f) &= -\frac{(\frac{x}{2})^3}{(\frac{f}{2}+3)(\frac{f}{2}+2)(\frac{f}{2}+1)}\Delta_2^1(F_x, f) + \Delta_6^1(F_x, f)\\ &= -\frac{4}{\sqrt{f\pi}} \exp\left(-\frac{z_\alpha^2}{2}\right)\{1 + O(f^{-1/2})\}.\end{aligned}$$

In summary, (A.25) is proved.

A.2.2.4 Proof of Proposition A.1.1 (on Page 145)

We prove Proposition A.1.1 based on the notation in Section A.2.2.2 and Lemma A.1.2, which is proved in Section A.2.2.3 above. Particularly, we write the left hand side of (A.21) as $\Delta_{2h}^v(F_x, f)$ below. By (A.25), we know (A.21) holds for $v = 1$ and $h \in \{1, 2, 3, 4\}$. We next prove (A.21) for $v \geq 2$ when $h \in \{1, 2, 3, 4\}$, respectively.

(Part I) Proof for $h = 1$. When $v = 2$, by (A.24), we have

$$\Delta_2^2(F_x, f) = -\frac{1}{\Gamma(\frac{f}{2}+2)}\left(\frac{x}{2}\right)^{\frac{f}{2}+1} e^{-x/2} + \frac{1}{\Gamma(\frac{f}{2}+1)}\left(\frac{x}{2}\right)^{\frac{f}{2}} e^{-x/2}.$$

Then we can write $\Delta_2^2(F_x, f) = A_1(f)Q_1(f)$, where we define

$$Q_1(f) = \Delta_2^1(F_x, f), \quad \text{and} \quad A_1(f) = x/(f+2) - 1. \tag{A.82}$$

Note that $Q_1(f) = O(f^{-1/2})$ by (A.25), and $A_1(f) = O(f^{-1/2})$ by (A.6) when $x = \chi_f^2(\alpha)$. Therefore, (A.21) holds for $h = 1$ and $v = 2$.

We next prove (A.21) for $h = 1$ and $v > 2$ by the mathematical induction. Assume that there exists some constant C such that uniformly for integers $1 \le k \le v-1$, $\Delta_2^k(F_x, f) = O(k!C^k f^{-k/2})$, that is, uniformly for integers $1 \le k \le v-1$,

$$\Delta_2^{k-1}(Q_1, f) = O(k!C^k f^{-k/2}). \tag{A.83}$$

We next prove $\Delta_2^v(F_x, f) = O(v!C^v f^{-v/2})$. By the definition of $Q_1(f)$ and $A_1(f)$, we have $\Delta_2^v(F_x, f) = \Delta_2^{v-1}(Q_1, f) = \Delta_2^{v-2}(A_1Q_1, f)$. By Lemma A.2.4,

$$\Delta_2^{v-2}(A_1Q_1, f) = \sum_{w=0}^{v-2} \binom{v-2}{w} \Delta_2^w(A_1, f)\Delta_2^{v-2-w}(Q_1, f+2w). \tag{A.84}$$

To evaluate (A.84), by (A.83), for $0 \le w \le v-2$, we have

$$\Delta_2^{v-2-w}(Q_1, f+2w) = O\big\{(v-w-1)!C^{v-w-1}f^{-(v-w-1)/2}\big\}.$$

In addition, to evaluate $\Delta_2^w(A_1, f)$ in (A.84), we use the following Lemma A.2.7.

Lemma A.2.7. *When $x = \chi_f^2(\alpha)$ and $f \to \infty$, $A_1(f) = \sqrt{2}z_\alpha f^{-1/2}\{1 + O(f^{-1})\}$, and for any integer $w \ge 1$,*

$$\Delta_2^w(A_1, f) = x \times (-1)^w 2^w w! \frac{1}{\prod_{k=1}^{w+1}(f+2k)}. \tag{A.85}$$

Thus there exists a constant C such that (A.85) *is of the order of $O(w!C^w f^{-w})$ as $f \to \infty$ uniformly for $w \ge 1$.*

Proof. Please see Section A.2.2.6 on Page 208. □

By Lemma A.2.7, (A.84) gives that as $f \to \infty$,

$$\begin{aligned}
&\Delta_2^{v-2}(A_1 Q_1, f) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A.86)}\\
&= O(f^{-1/2}) \times O\big\{(v-1)! C^{v-1} f^{-(v-1)/2}\big\}\\
&\quad + \sum_{w=1}^{v-2} \binom{v-2}{w} O\big(w! C^w f^{-w}\big) \times O\big\{(v-w-1)! C^{v-w-1} f^{-(v-w-1)/2}\big\}\\
&= (v-1)! C^{v-1} O(f^{-v/2}) + \sum_{w=1}^{v-2} (v-2)!(v-w-1) C^{v-1} O\{f^{-(w-1)/2} \times f^{-v/2}\}\\
&= O(v! C^v f^{-v/2}),
\end{aligned}$$

where in the last equation, we use $v - w - 1 \leq v - 1$ and $O\{f^{-(w-1)/2}\} = O(1)$ when $w \geq 1$. We note that there exists a constant C such that the last equation in (A.86) holds uniformly for $v \geq 1$. In summary, we obtain (A.21) for $h = 1$.

(Part II) Proof for $h = 2$. By (A.24), (A.81) and (A.82),

$$\Delta_4^1(F_x, f) = Q_2(f) + Q_1(f), \tag{A.87}$$

where we define

$$Q_2(f) = -\frac{1}{\Gamma(\frac{f}{2}+2)} \left(\frac{x}{2}\right)^{\frac{f}{2}+1} e^{-x/2}.$$

Then by (A.87) and Lemma A.2.5, we have

$$\Delta_4^v(F_x, f) = \Delta_4^{v-1}(Q_2, f) + \Delta_4^{v-1}(Q_1, f).$$

Therefore, to prove (A.21) for $h = 2$, it suffices to prove

$$\Delta_4^{v-1}(Q_1, f) = O(v! C^v f^{-v/2}),$$
$$\Delta_4^{v-1}(Q_2, f) = O(v! C^v f^{-v/2}). \tag{A.88}$$

As $Q_1(f) = Q_2(f-2)$, it suffices to prove (A.88), and we next use the mathematical induction. Note that (A.88) holds for $v = 1$ since $\Delta_4^0(Q_2, f) = Q_2(f) = O(f^{-1/2})$ by the proof of (A.25). In addition, for $v = 2$, we have

$$\Delta_4^1(Q_2, f) = Q_2(f+4) - Q_2(f) = A_2(f)Q_2(f), \tag{A.89}$$

where

$$A_2(f) = \frac{(\frac{x}{2})^2}{(\frac{f}{2}+3)(\frac{f}{2}+2)} - 1.$$

Note that $Q_2(f) = O(f^{-1/2})$, and when $x = \chi_f^2(\alpha)$, we have $A_2(f) = O(f^{-1/2})$ by (A.6). Therefore, $\Delta_4^1(Q_2, f) = O(f^{-1})$, i.e., (A.88) holds for $v = 2$. For $v \geq 3$, we next use the mathematical induction, where we assume for integers $0 \leq w \leq v-2$,

$$\Delta_4^w(Q_2, f) = O\{(w+1)! C^{w+1} f^{-(w+1)/2}\}, \tag{A.90}$$

and prove (A.88). By (A.89), $\Delta_4^{v-1}(Q_2, f) = \Delta_4^{v-2}(A_2 Q_2, f)$. Then by Lemma A.2.4,

$$\Delta_4^{v-2}(A_2 Q_2, f) = \sum_{w=0}^{v-2} \binom{v-2}{w} \Delta_4^w(A_2, f) \Delta_4^{v-2-w}(Q_2, f+4w). \tag{A.91}$$

We next prove (A.88) by (A.90), (A.91) and the following Lemma A.2.8.

Lemma A.2.8. *When $x = \chi_f^2(\alpha)$, $A_2(f) = 2\sqrt{2} z_\alpha f^{-1/2}\{1 + O(f^{-1/2})\}$. Moreover,* *there exists a constant C such that uniformly for any integer $w \geq 1$,*

$$\Delta_4^w(A_2, f) = O\Big\{(w+1)!C^w \prod_{t=1}^{w}(f+2t)^{-1}\Big\}. \tag{A.92}$$

Proof. Please see Section A.2.2.7 on Page 209. $\square$

By Lemma A.2.8 and (A.90), we have

$$\begin{aligned}
\text{(A.91)} =& O(f^{-1/2}) \times O\big\{(v-1)!C^{v-1}f^{-(v-1)/2}\big\} \\
&+ \sum_{w=1}^{v-2}\binom{v-2}{w} O\Big\{(w+1)!C^w \prod_{t=1}^{w}(f+2t)^{-1}(v-1-w)!C^{(v-1-w)}f^{-\frac{v-1-w}{2}}\Big\} \\
=& O\big\{(v-1)!C^{v-1}f^{-\frac{v}{2}}\big\} + \sum_{w=1}^{v-2} O\big\{(v-2)!(v-1-w)C^{v-1}f^{-\frac{v}{2}}\big\}\frac{(w+1)f^{\frac{w+1}{2}}}{\prod_{t=1}^{w}(f+2t)}.
\end{aligned} \tag{A.93}$$

To evaluate (A.93), we note that when $w = 1$ and 2, $(w+1)f^{(w+1)/2}\{\prod_{t=1}^{w}(f+2t)\}^{-1} = O(f^{(1-w)/2})$; when $w \geq 3$, as $f \to \infty$,

$$\frac{(w+1)f^{(w+1)/2}}{\prod_{t=1}^{w}(f+2t)} \leq \frac{w+1}{2w}\frac{f^{(w+1)/2}}{f^{w-1}} = O(1)$$

uniformly over $w \geq 3$. Moreover, by $\sum_{w=1}^{v-2}(v-2)!(v-1-w) \leq v!$, we obtain (A.91) $= O(v!C^v f^{-v/2})$.

(Part III) Proof for $h = 3$. By (A.24),

$$\Delta_6^1(F_x, f) = Q_3(f) + Q_2(f) + Q_1(f), \tag{A.94}$$

where we define $Q_3(f) = -\frac{1}{\Gamma(\frac{f}{2}+3)}\left(\frac{x}{2}\right)^{\frac{f}{2}+2} e^{-x/2}$. Then by (A.94) and Lemma A.2.5, $\Delta_6^v(F_x, f) = \Delta_6^{v-1}(Q_3, f) + \Delta_6^{v-1}(Q_2, f) + \Delta_6^{v-1}(Q_1, f)$. Since $Q_2(f) = Q_3(f-2)$ and

$Q_1(f) = Q_3(f-4)$, it suffices to prove

$$\Delta_6^{v-1}(Q_3, f) = O(v! C^v f^{-v/2}). \tag{A.95}$$

We next prove (A.95) by the mathematical induction. Note that (A.95) holds for $v = 1$ since $\Delta_6^0(Q_3, f) = Q_3(f) = O(f^{-1/2})$ by the proof of (A.25) in Section A.2.2.3. In addition, for $v = 2$,

$$\Delta_6^1(Q_3, f) = Q_3(f+6) - Q_3(f) = A_3(f) Q_3(f), \tag{A.96}$$

where $A_3(f) = \prod_{k=1}^3 A_{3,k}(f) - 1$ and $A_{3,k}(f) = x/(f + 4 + 2k)$. Note that $A_3(f) = O(f^{-1/2})$ when $x = \chi_f^2(\alpha)$ by (A.6). Moreover, as $Q_3(f) = O(f^{-1/2})$, $\Delta_6^1(Q_3, f) = O(f^{-1})$, i.e., (A.95) holds for $v = 2$. For $v \geq 3$, we next use the mathematical induction, where we assume for integers $0 \leq w \leq v - 2$,

$$\Delta_6^w(Q_3, f) = O\{(w+1)! C^{w+1} f^{-(w+1)/2}\}, \tag{A.97}$$

and prove (A.95). By (A.96), $\Delta_6^{v-1}(Q_3, f) = \Delta_6^{v-2}(A_3 Q_3, f)$. Then by Lemma A.2.4,

$$\Delta_6^{v-2}(A_3 Q_3, f) = \sum_{w=0}^{v-2} \binom{v-2}{w} \Delta_6^w(A_3, f) \Delta_6^{v-2-w}(Q_3, f + 6w). \tag{A.98}$$

We next prove (A.98) by (A.97) and the following Lemma A.2.9.

Lemma A.2.9. *When $x = \chi_f^2(\alpha)$, $A_3(f) = 3\sqrt{2} z_\alpha f^{-1/2}\{1 + O(f^{-1/2})\}$. Moreover, there exists a constant C such that uniformly for any integer $w \geq 1$,*

$$\Delta_6^w(A_3, f) = O\left\{(w+2)! C^w \prod_{t=1}^{w} (f + 2t)^{-1}\right\}.$$

Proof. Please see Section A.2.2.8 on Page 211. □

Then by (A.97) and Lemma A.2.9,

$$\begin{aligned}(\text{A.98}) &= O(f^{-1/2}) \times O\big\{(v-1)!C^{v-1}f^{-(v-1)/2}\big\} \\ &\quad + \sum_{w=1}^{v-2}\binom{v-2}{w} O\Big\{(w+2)!C^{w}\prod_{t=1}^{w}(f+2t)^{-1}(v-1-w)!C^{v-1-w}f^{-(v-1-w)/2}\Big\} \\ &= O\big\{(v-1)!C^{v-1}f^{-v/2}\big\} + \sum_{w=1}^{v-2} O\big\{(v-1)!C^{v}f^{-v/2}\big\}\frac{(w+2)(w+1)f^{(w+1)/2}}{\prod_{t=1}^{w}(f+2t)}.\end{aligned}$$

Note that when $w \le 4$, $(w+2)(w+1)f^{(w+1)/2}\prod_{t=1}^{w}(f+2t)^{-1} = O\{f^{(1-w)/2}\}$; when $w \ge 5$,

$$\frac{(w+2)(w+1)f^{(w+1)/2}}{\prod_{t=1}^{w}(f+2t)} \le \frac{(w+2)(w+1)}{w(w-1)}f^{(5-w)/2} = O(1)$$

as $f \to \infty$ uniformly over $w \ge 5$. It follows that (A.98) $= O(v!C^{v}f^{-v})$ and thus (A.95) is proved.

(Part IV) Proof for $h = 4$. By (A.24),

$$\Delta_8^1(F_x, f) = Q_4(f) + Q_3(f) + Q_2(f) + Q_1(f), \tag{A.99}$$

where we define $Q_4(f) = -\frac{1}{\Gamma(\frac{f}{2}+4)}\left(\frac{x}{2}\right)^{\frac{f}{2}+3}e^{-x/2}$. Then by (A.99) and Lemma A.2.4,

$$\Delta_8^v(F_x, f) = \Delta_8^{v-1}(Q_4, f) + \Delta_8^{v-1}(Q_3, f) + \Delta_8^{v-1}(Q_2, f) + \Delta_8^{v-1}(Q_1, f).$$

Since $Q_3(f) = Q_4(f-2)$, $Q_2(f) = Q_4(f-4)$, and $Q_1(f) = Q_4(f-6)$, it suffices to prove

$$\Delta_8^{v-1}(Q_4, f) = O(v!C^{v}f^{-v/2}). \tag{A.100}$$

We next prove (A.100) by the mathematical induction. Note that (A.100) holds for $v = 1$ since $\Delta_8^0(Q_4, f) = Q_4(f) = O(f^{-1/2})$ by the proof of (A.25) in Section A.2.2.3. In addition, for $v = 2$, we have

$$\Delta_8^1(Q_4, f) = Q_4(f+8) - Q_4(f) = A_4(f)Q_4(f), \tag{A.101}$$

where $A_4(f) = \prod_{k=1}^4 A_{4,k}(f) - 1$ and $A_{4,k}(f) = x/(f+6+2k)$. Note that $A_4(f) = O(f^{-1/2})$ as $x = f+\sqrt{2f}\{z_\alpha + O(f^{-1/2})\}$. Moreover, as $Q_4(f) = O(f^{-1/2})$, $\Delta_8^1(Q_4, f) = O(f^{-1})$, i.e., (A.100) holds for $v = 2$. For $v \geq 3$, we next use the mathematical induction, where we assume for integers $0 \leq w \leq v-2$,

$$\Delta_8^w(Q_4, f) = O\{(w+1)!C^{w+1}f^{-(w+1)/2}\}, \tag{A.102}$$

and prove (A.100). By (A.101), $\Delta_8^{v-1}(Q_4, f) = \Delta_8^{v-2}(A_4Q_4, f)$. Then by Lemma A.2.4,

$$\Delta_8^{v-2}(A_4Q_4, f) = \sum_{w=0}^{v-2} \binom{v-2}{w} \Delta_8^w(A_4, f)\Delta_8^{v-2-w}(Q_4, f+8w). \tag{A.103}$$

We next prove (A.103) by (A.102), (A.103) and the following Lemma A.2.10.

Lemma A.2.10. *When $x = \chi_f^2(\alpha)$, $A_4(f) = 4\sqrt{2}z_\alpha f^{-1/2}\{1+O(f^{-1/2})\}$. Moreover, there exists a constant C such that as $f \to \infty$,*

$$\Delta_8^w(A_4, f) = O\left\{(w+3)!C^w \prod_{t=1}^{w}(f+2t)^{-1}\right\}$$

holds uniformly for any integer $w \geq 1$

Proof. Please see Section A.2.2.9 on Page 211. $\square$

Then by (A.102) and Lemma A.2.10,

$$
\begin{aligned}
(\text{A.103}) &= O(f^{-1/2}) \times O\big\{(v-1)!C^{v-1}f^{-(v-1)/2}\big\} \\
&\quad + \sum_{w=1}^{v-2}\binom{v-2}{w} O\Big\{(w+3)!C^{w}\prod_{t=1}^{w}(f+2t)^{-1}(v-1-w)!C^{v-1-w}f^{-(v-1-w)/2}\Big\} \\
&= O\big\{(v-1)!C^{v-1}f^{-v/2}\big\} + \sum_{w=1}^{v-2} O\big\{(v-1)!C^{v}f^{-v/2}\big\}\frac{(w+3)(w+2)(w+1)f^{\frac{w+1}{2}}}{\prod_{t=1}^{w}(f+2t)}.
\end{aligned}
$$

Note that when $w \leq 6$, $(w+3)(w+2)(w+1)f^{(w+1)/2}\prod_{t=1}^{w}(f+2t)^{-1} = O\{f^{(1-w)/2}\}$; when $w \geq 7$, as $f \to \infty$,

$$
\frac{(w+3)(w+2)(w+1)f^{(w+1)/2}}{\prod_{t=1}^{w}(f+2t)} \leq \frac{(w+3)(w+2)(w+1)}{w(w-1)(w-2)}f^{(7-w)/2} = O(1)
$$

holds uniformly over $w \geq 7$. It follows that (A.103) $= O(v!C^{v}f^{-v})$ and thus (A.100) is proved.

A.2.2.5 Proof of Proposition A.1.2 (on Page 145)

Similar to the proof of Proposition A.1.1 in Section A.2.2.5, we prove Proposition A.1.2 using the notation in Section A.2.2.2 and Lemma A.1.2. We next discuss $(h_1, h_2) = (1, 2)$ and $(h_1, h_2) = (2, 3)$ in *(Part I)* and *(Part II)* below, respectively.

(Part I) Proof for $h_1 = 1$ and $h_2 = 2$. Based on the notation in Section A.2.2.2, it is equivalent to prove that there exists some constant C such that when $x = \chi^2_f(\alpha)$, as $f \to \infty$,

$$
\Delta_4^{v_2}\Delta_2^{v_1}(F_x, f) = O\big\{v_1!v_2!C^{v_1+v_2}f^{-(v_1+v_2)/2}\big\}, \tag{A.104}
$$

uniformly for integers $v_1, v_2 \geq 1$.

When $v_1 = 0$ or $v_2 = 0$, (A.104) holds by Proposition A.1.1. When $v_1 = v_2 = 1$, by (A.81), we have

$$\Delta_4^1\Delta_2^1(F_x,f) = -\frac{1}{\Gamma(\frac{f}{2}+3)}\left(\frac{x}{2}\right)^{\frac{f}{2}+2}e^{-x/2}+\frac{1}{\Gamma(\frac{f}{2}+1)}\left(\frac{x}{2}\right)^{\frac{f}{2}}e^{-x/2} = D_{2,4}(f)\Delta_2^1(F_x,f),$$

where

$$D_{2,4}(f) = \frac{x^2}{(f+4)(f+2)} - 1. \tag{A.105}$$

As $D_{2,4}(f) = O(f^{-1/2})$ and $\Delta_2^1(F_x,f) = O(f^{-1/2})$, (A.104) holds for $v_1 = v_2 = 1$. We next prove (A.104) by the mathematical induction. Particularly, we assume for integers $s_1 \le v_1$ and $s_2 \le v_2$,

$$\Delta_4^{s_2}\Delta_2^{s_1}(F_x,f) = O\big\{s_1!s_2!C^{s_1+s_2}f^{-(s_1+s_2)/2}\big\}, \tag{A.106}$$

and prove that (A.106) also holds for $(s_1,s_2) = (v_1+1,v_2)$ and $(s_1,s_2) = (v_1,v_2+1)$, i.e., $\Delta_4^{v_2}\Delta_2^{v_1+1}(F_x,f)$ and $\Delta_4^{v_2+1}\Delta_2^{v_1}(F_x,f)$, respectively.

Step I.1. $\Delta_4^{v_2}\Delta_2^{v_1+1}(F_x,f)$. Recall that we define $Q_1(f) = \Delta_2^1(F_x,f)$. It follows that (A.106) gives that for integers $s_1 \le v_1 - 1$ and $s_2 \le v_2$

$$\Delta_4^{s_2}\Delta_2^{s_1}(Q_1,f) = O\big\{(s_1+1)!s_2!C^{s_1+s_2+1}f^{-(s_1+s_2+1)/2}\big\}. \tag{A.107}$$

It is then equivalent to prove that (A.107) holds for $(s_1,s_2) = (v_1,v_2)$, that is, $\Delta_4^{v_2}\Delta_2^{v_1}(Q_1,f)$. By $\Delta_2^1(Q_1,f) = A_1(f)Q_1(f)$, (see the definitions in (A.82)), and

Lemmas A.2.4 and A.2.5,

$$
\begin{aligned}
&\Delta_4^{v_2}\Delta_2^{v_1}(Q_1, f) \\
&= \sum_{w_1=0}^{v_1-1}\binom{v_1-1}{w_1}\Delta_4^{v_2}\Big\{\Delta_2^{w_1}(A_1, f)\Delta_2^{v_1-1-w_1}(Q_1, f+2w_1)\Big\} \\
&= \sum_{w_1=0}^{v_1-1}\binom{v_1-1}{w_1}\sum_{w_2=0}^{v_2}\binom{v_2}{w_2}\Delta_4^{w_2}\Delta_2^{w_1}(A_1, f)\Delta_4^{v_2-w_2}\Delta_2^{v_1-1-w_1}(Q_1, f+2w_1+4w_2).
\end{aligned}
\tag{A.108}
$$

To evaluate (A.108), we use the following Lemma A.2.11.

Lemma A.2.11. *For two integers w_1 and w_2 satisfying $w_1 + w_2 \geq 1$, there exists some constant C such that as $f \to \infty$,*

$$
\Delta_4^{w_2}\Delta_2^{w_1}(A_1, f) = (w_1+w_2)!O\left(C^{w_1+w_2}\prod_{k=1}^{w_1+w_2}\frac{1}{f+2k}\right)
$$

uniformly over $w_1 + w_2 \geq 1$.

Proof. Please see Section A.2.2.10 on Page 212. □

By Lemma A.2.11 and the assumption (A.107), we have

$$
\begin{aligned}
(\text{A.108}) = &\sum_{w_1=0}^{v_1-1}\binom{v_1-1}{w_1}\sum_{w_2=0}^{v_2}\binom{v_2}{w_2}(w_1+w_2)!O\left(\prod_{k=1}^{w_1+w_2}\frac{1}{f+2k}\right)C^{v_1+v_2+1} \\
&\times (v_2-w_2)!(v_1-w_1)!O\big\{(f+2w_1+4w_2)^{-(v_1-w_1+v_2-w_2)/2}\big\} \\
= &\sum_{w_1=0}^{v_1-1}(v_1-1)!(v_1-w_1)\sum_{w_2=0}^{v_2}v_2!C^{v_1+v_2+1}O\{f^{-(v_1+v_2+1)/2}\} \\
&\times \frac{(w_1+w_2)!}{w_1!w_2!}\prod_{k=1}^{w_1+w_2}\frac{1}{f+2k} \times f^{(w_1+w_2+1)/2}.
\end{aligned}
$$

We next use the following Lemma A.2.12.

Lemma A.2.12. *For integers w_1, w_2, and f,*

$$\frac{(w_1+w_2)!}{w_1!w_2!}\prod_{k=1}^{w_1+w_2}\frac{1}{f+2k}\times f^{(w_1+w_2+1)/2}=O\big\{2^{-(w_1+w_2-1)/2}\big\}.$$

Proof. Please see Section A.2.2.11 on Page 214. $\square$

It follows that by Lemma A.2.12,

$$\begin{aligned}(\text{A.108}) &= \sum_{w_1=0}^{v_1-1}(v_1-1)!(v_1-w_1)\sum_{w_2=0}^{v_2}v_2!\frac{1}{(\sqrt{2})^{w_1+w_2-1}}C^{v_1+v_2+1}O\{f^{-(v_1+v_2+1)/2}\}\\ &= O\{v_1!v_2!C^{v_1+v_2+1}f^{-(v_1+v_2+1)/2}\},\end{aligned} \tag{A.109}$$

which is $O\{(v_1+1)!v_2!C^{v_1+v_2+1}f^{-(v_1+v_2+1)/2}\}$ as $v_1<v_1+1$. Therefore, we obtain $\Delta_4^{v_2}\Delta_2^{v_1}(Q_1,f)=O\{(v_1+1)!v_2!C^{v_1+v_2+1}f^{-(v_1+v_2+1)/2}\}$.

Step I.2. $\Delta_4^{v_2+1}\Delta_2^{v_1}(F_x,f)$. By (A.87),

$$\Delta_4^{v_2+1}\Delta_2^{v_1}(F_x,f)=\Delta_2^{v_1}\Delta_4^{v_2+1}(F_x,f)=\Delta_2^{v_1}\Delta_4^{v_2}(Q_2,f)+\Delta_4^{v_2}\Delta_2^{v_1+1}(F_x,f).$$

By (A.109), we have $\Delta_4^{v_2}\Delta_2^{v_1+1}(F_x,f)=O\{v_1!v_2!C^{v_1+v_2+1}f^{-(v_1+v_2+1)/2}\}$. Therefore, it remains to prove $\Delta_2^{v_1}\Delta_4^{v_2}(Q_2,f)=O\{v_1!(v_2+1)!C^{v_1+v_2+1}f^{-(v_1+v_2+1)/2}\}$. By (A.89) and Lemma A.2.4,

$$\begin{aligned}&\Delta_2^{v_1}\Delta_4^{v_2}(Q_2,f)\\ &=\Delta_2^{v_1}\left\{\sum_{w_2=0}^{v_2-1}\binom{v_2-1}{w_2}\Delta_4^{w_2}(A_2,f)\Delta_4^{v_2-1-w_2}(Q_2,f+4w_2)\right\}\\ &=\sum_{w_2=0}^{v_2-1}\binom{v_2-1}{w_2}\sum_{w_1=0}^{v_1}\binom{v_1}{w_1}\Delta_2^{w_1}\Delta_4^{w_2}(A_2,f)\Delta_2^{v_1-w_1}\Delta_4^{v_2-1-w_2}(Q_2,f+4w_2+2w_1).\end{aligned} \tag{A.110}$$

To evaluate (A.110) through the mathematical induction, by (A.87) and (A.106), we can assume that or integers $s_1 \le v_1$ and $s_2 \le v_2 - 1$,

$$\Delta_2^{s_1} \Delta_4^{s_2}(Q_2, f) = O\big\{s_1!(s_2+1)!C^{s_1+s_2+1} f^{-(s_1+s_2+1)/2}\big\}. \tag{A.111}$$

In addition, we use the following Lemma A.2.13.

Lemma A.2.13. *For two integers w_1 and w_2 satisfying $w_1 + w_2 \ge 1$,*

$$\Delta_4^{w_2} \Delta_2^{w_1}(A_2, f) = (w_1 + w_2 + 1)! O\left(C^{w_1+w_2+1} \prod_{k=1}^{w_1+w_2} \frac{1}{f+2k} \right).$$

Proof. Please see Section A.2.2.12 on Page 216. □

Combining (A.111) and Lemma A.2.13, we obtain $\Delta_2^{v_1}\Delta_4^{v_2}(Q_2, f) = O\{v_1!v_2!C^{v_1+v_2+1} \times f^{-(v_1+v_2+1)/2}\}$ similarly to (A.109) in *Step I.1.* As $v_2 < v_2+1$, we have $\Delta_2^{v_1}\Delta_4^{v_2}(Q_2, f) = O\{v_1!(v_2+1)!C^{v_1+v_2+1} f^{-(v_1+v_2+1)/2}\}$.

(Part II) Proof for $h_1 = 2$ and $h_2 = 3$. In this part, we prove

$$\Delta_6^{v_2} \Delta_4^{v_1}(F_x, f) = O\big\{v_1!v_2!C^{v_1+v_2} f^{-(v_1+v_2)/2}\big\}, \tag{A.112}$$

as $f \to \infty$ and uniformly for integers $v_1, v_2 \ge 1$.

When $v_1 = 0$ or $v_2 = 0$, (A.112) holds by Proposition A.1.1. When $v_1 = v_2 = 1$, note that $\Delta_4^1(F_x, f) = Q_1(f) + Q_2(f)$ by (A.87). Then we have $\Delta_6^1\Delta_4^1(F_x, f) = \Delta_6^1(Q_1, f) + \Delta_6^1(Q_2, f)$. Particularly,

$$\begin{aligned}
\Delta_6^1(Q_1, f) &= D_{2,6}(f)Q_1(f), \qquad D_{2,6}(f) = \frac{x^3}{(f+6)(f+4)(f+2)} - 1; \\
\Delta_6^1(Q_2, f) &= D_{4,6}(f)Q_2(f), \qquad D_{4,6}(f) = \frac{x^3}{(f+8)(f+6)(f+4)} - 1.
\end{aligned} \tag{A.113}$$

By the proof of (A.25), $Q_1(f) = O(f^{-1/2})$ and $Q_2(f) = O(f^{-1/2})$. In addition, for $x = \chi_f^2(\alpha)$, by (A.6), $D_{2,6}(f) = O(f^{-1/2})$ and $D_{4,6}(f) = O(f^{-1/2})$. Therefore, (A.112) holds for $v_1 = 1$ and $v_2 = 1$. When $v_1 > 1$ or $v_2 > 1$, by (A.87),

$$\Delta_6^{v_2}\Delta_4^{v_1}(F_x, f) = \Delta_6^{v_2}\Delta_4^{v_1-1}(Q_1, f) + \Delta_6^{v_2}\Delta_4^{v_1-1}(Q_2, f).$$

It suffices to prove

$$\Delta_6^{v_2}\Delta_4^{v_1-1}(Q_1, f) = O\big\{v_1!v_2!C^{v_1+v_2}f^{-(v_1+v_2)/2}\big\}, \tag{A.114}$$

$$\Delta_6^{v_2}\Delta_4^{v_1-1}(Q_2, f) = O\big\{v_1!v_2!C^{v_1+v_2}f^{-(v_1+v_2)/2}\big\}. \tag{A.115}$$

We next prove (A.114) and (A.115) by the mathematical induction, respectively.

First, to prove (A.114), we apply the mathematical induction considering increasing v_1 and v_2 in the following *Step II.1* and *Step II.2*, respectively.

Step II.1. We assume for $0 \le s_1 \le v_1 - 2$ and $0 \le s_2 \le v_2$,

$$\Delta_6^{s_2}\Delta_4^{s_1}(Q_1, f) = O\{(s_1+1)!s_2!C^{s_1+s_2+1}f^{-(s_1+s_2+1)/2}\}, \tag{A.116}$$

and then prove (A.114). Note that $\Delta_4^1(Q_1, f) = D_{2,4}(f)Q_1(f)$, where $D_{2,4}(f)$ is defined in (A.105). Then by the Leibniz rule in Lemma A.2.4,

$$\begin{aligned}\Delta_6^{v_2}\Delta_4^{v_1-1}(Q_1, f) &= \Delta_6^{v_2}\Delta_4^{v_1-2}(D_{2,4}Q_1, f)\\ &= \sum_{k_2=0}^{v_2}\sum_{k_1=0}^{v_1-2}\binom{v_1-2}{k_1}\binom{v_2}{k_2}\Delta_6^{k_2}\Delta_4^{k_1}(D_{2,4}, f)\\ &\quad\times \Delta_6^{v_2-k_2}\Delta_4^{v_1-2-k_1}(Q_1, f+4k_1+6k_2).\end{aligned} \tag{A.117}$$

To evaluate (A.117), we use the following Lemma A.2.14.

Lemma A.2.14. *For integers $k_1 + k_2 \geq 1$, there exists some constant C such that*

$$\Delta_6^{k_2} \Delta_4^{k_1}(D_{2,4}, f) = (k_1 + k_2 + 1)! O\left(C^{k_1+k_2} \prod_{t=1}^{k_1+k_2} \frac{1}{f + 2t}\right),$$

as $f \to \infty$ and uniformly over $k_1 + k_2 \geq 1$.

Proof. Please see Section A.2.2.13 on Page 217. □

Then applying similar analysis to that of (A.108) and (A.109) in *Part I* above, we obtain (A.114) by the assumption (A.116) and Lemma A.2.14.

Step II.2. We assume for $0 \leq s_1 \leq v_1 - 1$ and $0 \leq s_2 \leq v_2 - 1$, (A.116) holds, and then prove (A.114). By (A.113) and the Leibniz rule in Lemma A.2.4,

$$\begin{aligned} &\Delta_6^{v_2} \Delta_4^{v_1 - 1}(Q_1, f) \qquad \text{(A.118)} \\ &= \Delta_4^{v_1-1} \Delta_6^{v_2-1}(D_{2,6} Q_1, f) \\ &= \sum_{k_1=0}^{v_1-1} \sum_{k_2=0}^{v_2-1} \binom{v_1 - 1}{k_1} \binom{v_2 - 1}{k_2} \Delta_6^{k_2} \Delta_4^{k_1}(D_{2,6}, f) \times \Delta_6^{v_2-1-k_2} \Delta_4^{v_1-1-k_1}(Q_1, f + 4k_1 + 6k_2). \end{aligned}$$

Similarly to the analysis of (A.117), we use the following Lemma A.2.15 to evaluate (A.118).

Lemma A.2.15. *For integers $k_1 + k_2 \geq 1$, there exists a constant C such that*

$$\Delta_6^{k_2} \Delta_4^{k_1}(D_{2,6}, f) = (k_1 + k_2 + 2)! O\left(C^{k_1+k_2} \prod_{t=1}^{k_1+k_2} \frac{1}{f + 2t}\right),$$

as $f \to \infty$ and uniformly over $k_1 + k_2 \geq 1$.

Proof. Please see Section A.2.2.14 on Page 218. □

Since we assume (A.116) holds for $0 \leq s_1 \leq v_1 - 1$ and $0 \leq s_2 \leq v_2 - 1$, then by

Lemma A.2.15,

$$
\begin{aligned}
(\text{A.118}) = & \sum_{k_1=0}^{v_1-1}\sum_{k_2=0}^{v_2-1}\binom{v_1-1}{k_1}\binom{v_2-1}{k_2}(k_1+k_2+2)!(v_2-1-k_2)!(v_1-k_1)! \\
& \times C^{v_1+v_2}f^{-(v_1+v_2)/2}O\left(f^{-(k_1+k_2+1)/2}\prod_{t=1}^{k_1+k_2}\frac{1}{f+2t}\right) \\
= & \, C^{v_1+v_2}f^{-(v_1+v_2)/2}(v_1-1)!(v_2-1)!\sum_{k_2=0}^{v_2}\sum_{k_1=0}^{v_1-1}(v_1-k_1) \\
& \times \frac{(k_1+k_2+2)!}{k_1!k_2!}O\left(f^{-(k_1+k_2+1)/2}\prod_{t=1}^{k_1+k_2}\frac{1}{f+2t}\right).
\end{aligned}
$$

We next use the following Lemma A.2.16 to evaluate (A.118).

Lemma A.2.16. *For integers $k_1+k_2 \geq 1$, as $f \to \infty$,*

$$
\frac{(k_1+k_2+2)!}{k_1!k_2!}O\left\{f^{-(k_1+k_2+1)/2}\prod_{t=1}^{k_1+k_2}\frac{1}{f+2t}\right\} = O\{2^{-(k_1+k_2-1)/2}\}.
$$

Proof. Please see Section A.2.2.15 on Page 219. □

Then by Lemma A.2.16, we obtain $\Delta_6^{v_2}\Delta_4^{v_1-1}(Q_1, f) = O\{v_1!v_2!C^{v_1+v_2}f^{-(v_1+v_2)/2}\}$ similarly to (A.109). In summary, combining *Step II.1* and *Step II.2*, we finish the proof of (A.114).

Second, to prove (A.115), we can use the mathematical induction similarly to the proof of (A.114). The analysis would be very similar and the details are thus skipped.

A.2.2.6 Proof of Lemma A.2.7 (on Page 194)

When $x = \chi_f^2(\alpha)$, by (A.6), we have $x = f + \sqrt{2f}\{z_\alpha + O(f^{-1/2})\}$, and then $A_1(f) = \sqrt{2}z_\alpha f^{-1/2}\{1+O(f^{-1})\}$. We next prove (A.85) by the mathematical induction. For $w = 1$, we compute

$$\Delta_2^1(A_1, f) = A_1(f+2) - A_1(f) = -x \times 2 \times \frac{1}{(f+2)(f+4)}.$$

Therefore (A.85) holds when $w = 1$. We next assume (A.85) holds, and prove the conclusion holds for $\Delta_2^{w+1}(A_1, f)$. Particularly,

$$\begin{aligned}\Delta_2^{w+1}(A_1, f) &= x \times (-1)^w 2^w w! \left\{ \frac{1}{\prod_{k=2}^{w+2}(f+2k)} - \frac{1}{\prod_{k=1}^{w+1}(f+2k)} \right\} \\ &= x \times (-1)^{w+1} 2^{w+1} (w+1)! \frac{1}{\prod_{k=1}^{w+2}(f+2k)}.\end{aligned}$$

In summary, Lemma A.2.7 is proved.

A.2.2.7 Proof of Lemma A.2.8 (on Page 196)

When $x = \chi_f^2(\alpha)$, by (A.6), we have $x = f + \sqrt{2f}\{z_\alpha + O(f^{-1/2})\}$, and then $A_2(f) = 2\sqrt{2} z_\alpha f^{-1/2}\{1 + O(f^{-1})\}$. We next prove (A.92). Note that we can write $A_2(f) = A_{2,1}(f) A_{2,2}(f) - 1$, where we define

$$A_{2,1}(f) = \frac{x}{f+4} \quad \text{and} \quad A_{2,2}(f) = \frac{x}{f+6}.$$

By Lemmas A.2.4 and A.2.5, when $w \geq 1$,

$$\Delta_4^w(A_2, f) = \Delta_4^w(A_{2,1} A_{2,2}, f) = \sum_{k=0}^{w} \binom{w}{k} \Delta_4^k(A_{2,1}, f) \Delta_4^{w-k}(A_{2,2}, f+4k). \tag{A.119}$$

To prove (A.119) $= O(w! C^w f^{-w})$, we next evaluate $\Delta_4^k(A_{2,1}, f)$ and $\Delta_4^{w-k}(A_{2,2}, f + 4k)$.

In particular, we prove that

$$\Delta_4^k(A_{2,1}, f) = (-1)^k 4^k k! x \times \frac{1}{\prod_{t=1}^{k+1}(f+4t)} \tag{A.120}$$

by the mathematical induction. When $k = 1$,

$$\Delta_4^1(A_{2,1}, f) = \frac{x}{f+8} - \frac{x}{f+4} = \frac{x \times (-4)}{(f+4)(f+8)}.$$

Thus (A.120) holds for $k = 1$. We next assume (A.120) holds and prove the conclusion for $\Delta_4^{k+1}(A_{2,1}, f)$. Specifically,

$$\begin{aligned}\Delta_4^{k+1}(A_{2,1}, f) &= (-1)^k 4^k k! x \left\{ \frac{1}{\prod_{t=2}^{k+2}(f+4t)} - \frac{1}{\prod_{t=1}^{k+1}(f+4t)} \right\} \\ &= (-1)^{k+1} 4^{k+1} (k+1)! x \frac{1}{\prod_{t=1}^{k+2}(f+4t)}.\end{aligned}$$

In summary, (A.120) is proved. Moreover, as $A_{2,2}(f) = A_{2,1}(f+2)$, we have

$$\Delta_4^k(A_{2,2}, f) = \Delta_4^k(A_{2,1}, f+2) = (-1)^k 4^k k! x \frac{1}{\prod_{t=1}^{k+1}(f+2+4t)}.$$

It follows that $\Delta_4^{w-k}(A_{2,2}, f+4k) = (-1)^{w-k} 4^{w-k} (w-k)! x \{\prod_{t=k+1}^{w+1}(f+2+4t)\}^{-1}$. Then by (A.119), there exists a constant C such that

$$\begin{aligned}\left|\Delta_4^w(A_{2,1}A_{2,2}, f)\right| &= \left| \sum_{k=0}^{w} \binom{w}{k} \frac{(-4)^w k! (w-k)! x^2}{\prod_{t=1}^{k+1}(f+4t) \prod_{t=k+1}^{w+1}(f+2+4t)} \right| \\ &\leq w! C^w \sum_{k=0}^{w} \frac{x^2}{\prod_{t=1}^{w+2}(f+2t)}.\end{aligned}$$

As $x = \chi_f^2(\alpha) = O(f)$, we obtain that (A.92) holds as $f \to \infty$ and uniformly for any integer $w \geq 1$.

A.2.2.8 Proof of Lemma A.2.9 (on Page 198)

When $x = \chi_f^2(\alpha)$, by (A.6), we have $x = f + \sqrt{2f}\{z_\alpha + O(f^{-1/2})\}$, and then $A_3(f) = 3\sqrt{2}z_\alpha f^{-1/2}\{1 + O(f^{-1/2})\}$. We next consider $\Delta_6^w(A_3, f)$ for $w \geq 1$. As $A_3(f) = \prod_{l=1}^3 A_{3,l}(f) - 1$,

$$\Delta_6^w(A_3, f) = \sum_{k_1=0}^{w}\sum_{k_2=0}^{k_1}\binom{k_1}{k_2}\binom{w}{k_1}\Delta_6^{k_2}(A_{3,1}, f)\Delta_6^{k_1-k_2}(A_{3,2}, f+6k_2)\Delta_6^{w-k_1}(A_{3,3}, f+6k_1).$$

Similarly to the proofs of Lemma A.2.7 in Section A.2.2.6, for $A_{3,l}(f)$, $l \in \{1,2,3\}$, we can obtain that for any integer $w \geq 1$ and $l \in \{1,2,3\}$

$$\Delta_6^w(A_{3,l}, f) = (-6)^w w! x \times \frac{1}{\prod_{t=0}^{w}(f+4+2l+6t)}.$$

It follows that

$$\begin{aligned}\Delta_6^w(A_3, f) = &\sum_{k_1=0}^{w}\sum_{k_2=0}^{k_1}\binom{k_1}{k_2}\binom{w}{k_1}(-6)^w k_2!(k_1-k_2)!(w-k_1)! \\ &\times x^3\left\{\prod_{t=1}^{k_2+1}(f+6t)\prod_{t=k_2+1}^{k_1+1}(f+6t+2)\prod_{t=k_1+1}^{w+1}(f+6t+4)\right\}^{-1}.\end{aligned}$$

As $\binom{k_1}{k_2}\binom{w}{k_1}k_2!(k_1-k_2)!(w-k_1)! = w!$, $\sum_{k_1=0}^{w}\sum_{k_2=0}^{k_1} 1 \leq (w+1)^2$, and $x = \chi_f^2(\alpha) = O(f)$, there exists a constant C such that as $f \to \infty$ and uniformly over $w \geq 1$,

$$\Delta_6^w(A_3, f) = O\left\{(w+2)!C^w\prod_{t=1}^{w}(f+2t)^{-1}\right\}.$$

A.2.2.9 Proof of Lemma A.2.10 (on Page 200)

When $x = \chi_f^2(\alpha)$, by (A.6), we have $x = f + \sqrt{2f}\{z_\alpha + O(f^{-1/2})\}$, and then $A_4(f) = 4\sqrt{2}z_\alpha f^{-1/2}\{1 + O(f^{-1/2})\}$. We next prove the conclusion for $w \geq 1$. As

$A_4(f) = \prod_{l=1}^{4} A_{4,l}(f) - 1$,

$$\begin{aligned}\Delta_8^w(A_4, f) = &\sum_{k_1=0}^{w}\sum_{k_2=0}^{k_1}\sum_{k_3=0}^{k_2}\binom{w}{k_1}\binom{k_1}{k_2}\binom{k_2}{k_3}\Delta_8^{k_3}(A_{4,1}, f) \times \Delta_8^{k_2-k_3}(A_{4,2}, f+8k_3)\\ &\times \Delta_8^{k_1-k_2}(A_{4,3}, f+8k_2) \times \Delta_8^{w-k_1}(A_{4,4}, f+8k_1).\end{aligned}$$

Similarly to the proof of Lemma A.2.7 in Section A.2.2.6, for $A_{4,l}(f)$, $l \in \{1,2,3,4\}$, we can obtain that for any integer $w \geq 1$,

$$\Delta_8^w(A_{4,l}, f) = (-8)^w w! x \times \frac{1}{\prod_{t=0}^{w}(f+6+2l+8t)}.$$

It follows that

$$\begin{aligned}\Delta_8^w(A_4, f) = &\sum_{k_1=0}^{w}\sum_{k_2=0}^{k_1}\sum_{k_3=0}^{k_2}\binom{w}{k_1}\binom{k_1}{k_2}\binom{k_2}{k_3}(-8)^w k_3!(k_2-k_3)!(k_1-k_2)!(w-k_1)!\\ &\times x^4\left\{\prod_{t=1}^{k_3+1}(f+8t)\prod_{t=k_3+1}^{k_2+1}(f+8t+2)\prod_{t=k_2+1}^{k_1+1}(f+8t+4)\prod_{k_1+1}^{w+1}(f+8t+6)\right\}^{-1}.\end{aligned}$$

As $\binom{w}{k_1}\binom{k_1}{k_2}\binom{k_2}{k_3}k_3!(k_2-k_3)!(k_1-k_2)!(w-k_1)! = w!$, $\sum_{k_1=0}^{w}\sum_{k_2=0}^{k_1}\sum_{k_3=0}^{k_2} 1 \leq (w+1)^3$, and $x = O(f)$, there exists a constant C such that

$$\Delta_8^w(A_4, f) = O\left\{(w+3)!C^w\prod_{t=1}^{w}(f+2t)^{-1}\right\}.$$

A.2.2.10 Proof of Lemma A.2.11 (on Page 203)

By the proof of Lemma A.2.7, we have $\Delta_2^{w_1}(A_1, f) = (-1)^{w_1}2^{w_1}w_1!x\prod_{s=1}^{w_1+1}A_{1,s}(f)$, where $A_{1,s}(f) = 1/(f+2s)$. It follows that

$$\Delta_4^{w_2}\{\Delta_2^{w_1}(A_1, f)\} = x(-2)^{w_1}w_1!\Delta_4^{w_2}\left\{\prod_{s=1}^{w_1+1}A_{1,s}(f)\right\}. \tag{A.121}$$

To prove Lemma A.2.11, by $x = \chi_f^2(\alpha) = O(f)$ and (A.121), it suffices to prove

$$\Delta_4^{w_2}\left\{\prod_{s=1}^{w_1+1} A_{1,s}(f)\right\} = \frac{(w_1+w_2)!}{w_1!} O\left\{C^{w_1+w_2}\prod_{s=1}^{w_1+w_2+1}(f+2s)^{-1}\right\}. \qquad \text{(A.122)}$$

We next prove (A.122) by the mathematical induction. Consider $w_1 = 0$ first. Similarly to the proof of Lemma A.2.7, for each integer $1 \le s \le w_1 + 1$, we have

$$\Delta_4^{w_2}(A_{1,s}, f) = w_2!(-4)^{w_2}\prod_{k=0}^{w_2}(f+2s+4k). \qquad \text{(A.123)}$$

Thus (A.122) holds for $w_1 = 0$. We then assume for integers $1 \le l \le w_1$,

$$\Delta_4^{w_2}\left\{\prod_{s=1}^{l} A_{1,s}(f)\right\} = \frac{(w_2+l-1)!}{(l-1)!} O\left\{\prod_{k=1}^{w_2+l}(f+2k)^{-1}\right\}, \qquad \text{(A.124)}$$

and prove (A.122). By the Leibniz rule in Lemma A.2.4,

$$\Delta_4^{w_2}\left\{\prod_{s=1}^{w_1+1} A_{1,s}(f)\right\} = \sum_{k_2=0}^{w_2}\binom{w_2}{k_2}\Delta_4^{k_2}\left\{\prod_{s_1=1}^{w_1} A_{1,s_1}(f)\right\}\Delta_4^{w_2-k_2}(A_{1,w_1+1}, f+4k_2). \qquad \text{(A.125)}$$

Then by (A.123) and (A.124), we obtain

$$\begin{aligned}
\text{(A.125)} = {} & \sum_{k_2=0}^{w_2}\binom{w_2}{k_2}\frac{(k_2+w_1-1)!}{(w_1-1)!} O\left(C^{w_1+k_2}\prod_{s_1=1}^{w_1+k_2}\frac{1}{f+2s_1}\right) \\
& \times O\left\{(w_2-k_2)!C^{w_2-k_2}\prod_{s_2=0}^{w_2-k_2}\frac{1}{f+4k_2+2(w_1+1)+4s_2}\right\} \\
= {} & C^{w_1+w_2}\sum_{k_2=0}^{w_2} w_2!\binom{k_2+w_1-1}{k_2} O\left(\prod_{s=1}^{w_1+w_2+1}\frac{1}{f+2s}\right).
\end{aligned}$$

By the hockey-stick identity, $\sum_{k_2=0}^{w_2}\binom{k_2+w_1-1}{k_2} = \binom{w_1+w_2}{w_2}$. Therefore, (A.122) is proved and then (A.121) follows.

A.2.2.11 Proof of Lemma A.2.12 (on Page 203)

We next prove Lemma A.2.12 by discussing the cases when $w_1 + w_2$ is odd and even, respectively.

(1) When $w_1 + w_2$ is odd, $(w_1 + w_2 + 1)/2$ is an integer, and then

$$
\begin{aligned}
(w_1 + w_2)! \prod_{k=1}^{w_1+w_2} \frac{1}{f + 2k} \times f^{(w_1+w_2+1)/2} &\le (w_1 + w_2)! \prod_{k=(w_1+w_2+1)/2+1}^{w_1+w_2} \frac{1}{2k} \\
&\le 2^{-(w_1+w_2-1)/2} \prod_{k=1}^{(w_1+w_2+1)/2} k.
\end{aligned}
$$

To prove Lemma A.2.12, it now suffices to prove that there exists a constant C such that

$$
\frac{1}{w_1! w_2!} \prod_{k=1}^{(w_1+w_2+1)/2} k \le C. \tag{A.126}
$$

To prove (A.126), we use the following Lemma A.2.17.

Lemma A.2.17 (Factorial bound)**.** *For any integer $w \ge 1$,*

$$
\left(\frac{w}{e}\right)^w e \le w! \le \left(\frac{w+1}{e}\right)^{w+1} e.
$$

Proof. This is a known bound on factorial in literature, and is obtained by $\int_1^w \ln x \mathrm{d}x \le \sum_{x=1}^w \ln x \le \int_0^w \ln(x+1)\mathrm{d}x$. □

Assume without loss of generality that $w_2 \geq w_1$, and then by Lemma A.2.17,

$$\begin{aligned}
&\frac{1}{w_1!w_2!}\prod_{k=1}^{(w_1+w_2+1)/2} k \\
\leq&\frac{1}{e}\left(\frac{e}{w_1}\right)^{w_1}\left(\frac{e}{w_2}\right)^{w_2}\left(\frac{w_1+w_2+3}{2e}\right)^{(w_1+w_2+3)/2} \\
=&\frac{1}{e}\left(\frac{e^2}{w_1w_2}\frac{w_1+w_2+3}{2e}\right)^{w_1}\left(\frac{e^2}{w_2^2}\frac{w_1+w_2+3}{2e}\right)^{(w_2-w_1)/2}\left(\frac{w_1+w_2+3}{2e}\right)^{3/2}.
\end{aligned} \tag{A.127}$$

As $w_1 + w_2 + 3 \leq 4w_2$, there exists a constant C such that

$$(\text{A.127}) \leq C\left(\frac{2e}{w_1}\right)^{w_1}\left(\frac{2e}{w_2}\right)^{(w_2-w_1)/2}(w_1+w_2+3)^{3/2}.$$

When $w_2 - w_1 \geq 3$,

$$(\text{A.127}) \leq C\left(\frac{2e}{w_1}\right)^{w_1}\left(\frac{2e}{w_2}\right)^{(w_2-w_1-3)/2}\left\{\frac{2e(w_1+w_2+3)}{w_2}\right\}^{3/2},$$

which is bounded. When $0 \leq w_2 - w_1 \leq 2$, (A.127) $\leq C(2e/w_1)^{w_1}(2w_1+5)^{3/2}$, which is also bounded. In summary, (A.127) is bounded.

(2) When $w_1 + w_2$ is even, similarly, we have

$$(w_1+w_2)!\prod_{k=1}^{w_1+w_2}\frac{1}{f+2k}\times f^{(w_1+w_2+1)/2} \leq 2^{-(w_1+w_2)/2-1}\prod_{k=1}^{(w_1+w_2)/2+1} k.$$

To prove Lemma A.2.12, it now suffices to prove that there exists a constant C such that $\frac{1}{w_1!w_2!}\prod_{k=1}^{(w_1+w_2)/2+1} k \leq C$. Similar analysis can be applied and the conclusions follow.

A.2.2.12 Proof of Lemma A.2.13 (on Page 205)

When $w_1 = 0$, we know Lemma A.2.13 holds by Lemma A.2.8. Recall that we write $A_2(f) = A_{2,1}(f)A_{2,2}(f) - 1$ in Section A.2.2.7. Thus when $w_1 + w_2 \geq 1$,

$$\Delta_2^{w_1}\Delta_4^{w_2}(A_2, f) = \Delta_2^{w_1}\Delta_4^{w_2}(A_{2,1}A_{2,2}, f).$$

By the Leibniz rule in Lemma A.2.4,

$$\begin{aligned}&\Delta_2^{w_1}\Delta_4^{w_2}(A_{2,1}A_{2,2}, f)\\ =\ &\sum_{k_1=0}^{w_1}\sum_{k_2=0}^{w_2}\binom{w_1}{k_1}\binom{w_2}{k_2}\Delta_2^{k_1}\Delta_4^{k_2}(A_{2,1}, f)\Delta_2^{w_1-k_1}\Delta_4^{w_2-k_2}(A_{2,2}, f + 2k_1 + 4k_2).\end{aligned} \tag{A.128}$$

Following the proof of Lemma A.2.11, we have when $k_1 + k_2 \geq 1$,

$$\Delta_2^{k_1}\Delta_4^{k_2}(A_{2,1}, f) = (k_1 + k_2)!O\left(C^{k_1+k_2}\prod_{s=1}^{k_1+k_2}\frac{1}{f + 2s}\right),$$

and when $w_1 + w_2 - k_1 - k_2 \geq 1$,

$$\begin{aligned}&\Delta_2^{w_1-k_1}\Delta_4^{w_2-k_2}(A_{2,2}, f + 2k_1 + 4k_2)\\ =\ &(w_1 + w_2 - k_1 - k_2)!O\left(C^{w_1+w_2-k_1-k_2}\prod_{s=k_1+k_2+1}^{w_1+w_2}\frac{1}{f + 2s}\right).\end{aligned}$$

Therefore,

$$(\text{A.128}) = w_1!w_2!\sum_{k_1=0}^{w_1}\sum_{k_2=0}^{w_2}\binom{k_1 + k_2}{k_1}\binom{w_1 + w_2 - k_1 - k_2}{w_1 - k_1}O\left(C^{w_1+w_2}\prod_{s=1}^{w_1+w_2}\frac{1}{f + 2s}\right).$$

By the Chu–Vandermonde identity,

$$\sum_{k_1=0}^{w_1}\sum_{k_2=0}^{w_2}\binom{k_1+k_2}{k_1}\binom{w_1+w_2-k_1-k_2}{w_1-k_1} = \sum_{m=0}^{w_1+w_2}\sum_{s_1=0}^{w_1}\binom{m}{s_1}\binom{w_1+w_2-m}{w_1-s_1}$$
$$= (w_1+w_2+1)\binom{w_1+w_2}{w_1}.$$

Then $\Delta_2^{w_1}\Delta_4^{w_2}(A_2, f) = (w_1+w_2+1)!O\{C^{w_1+w_2}\prod_{s=1}^{w_1+w_2}(f+2s)^{-1}\}$.

A.2.2.13 Proof of Lemma A.2.14 (on Page 206)

By the definition of $D_{2,4}(f)$, when $k_1+k_2 \geq 1$,

$$\Delta_6^{k_2}\Delta_4^{k_1}(D_{2,4}, f) = x^2\Delta_6^{k_2}\Delta_4^{k_1}(A_{1,1}A_{1,2}, f),$$

where recall that we define $A_{1,t} = 1/(f+2t)$ for integers t. By the Leibniz rule in Lemma A.2.4,

$$\begin{aligned}
&\Delta_6^{k_2}\Delta_4^{k_1}(A_{1,1}A_{1,2}, f)\\
=&\sum_{s_2=0}^{k_2}\sum_{s_1=0}^{k_1}\binom{k_1}{s_1}\binom{k_2}{s_2}\Delta_6^{s_2}\Delta_4^{s_1}(A_{1,1}, f)\Delta_6^{k_2-s_2}\Delta_4^{k_1-s_1}(A_{1,2}, f+4s_1+6k_2)
\end{aligned}$$

Following the proof of Lemma A.2.11 in Section A.2.2.10, we similarly have

$$\Delta_6^{s_2}\Delta_4^{s_1}(A_{1,1}, f) = (s_1+s_2)!O\left(C^{s_1+s_2}\prod_{k=1}^{s_1+s_2+1}\frac{1}{f+2k}\right).$$

Then following the proof of Lemma A.2.13 in Section A.2.2.12, we obtain Lemma A.2.14. The analysis will be very similar and thus the details are skipped.

A.2.2.14 Proof of Lemma A.2.15 (on Page 207)

Note that we can write $D_{2,6}(f) = x^3 \prod_{k=1}^{3} A_{1,k}(f) - 1$. By the Leibniz rule in Lemma A.2.4,

$$\Delta_4^{k_1}\Delta_6^{k_2}(D_{2,6}, f) = \sum_{s_1=0}^{k_2}\sum_{s_2=0}^{s_1}\binom{k_1}{s_1}\binom{s_1}{s_2}\sum_{t_1=0}^{k_1}\sum_{t_2=0}^{t_1}\binom{k_2}{t_1}\binom{t_1}{t_2}x^3 \times \Delta_4^{t_2}\Delta_6^{s_2}(A_{3,1}, f)$$
$$\times \Delta_4^{t_1-t_2}\Delta_6^{s_1-s_2}(A_{3,2}, f + 6s_2 + 4t_2)\Delta_4^{k_1-t_1}\Delta_6^{k_2-s_1}(A_{3,3}, f + 6s_1 + 4t_1).$$

Following the proof of Lemma A.2.11 in Section A.2.2.10, we similarly have that for integers $t + s \geq 1$, and $l \in \{1, 2, 3\}$,

$$\Delta_4^t\Delta_6^s(A_{3,l}) = (t + s)!O\left(C^{t+s}\prod_{m=1}^{t+s+1}\frac{1}{f + 2m}\right).$$

By $x = \chi_f^2(\alpha) = O(f)$,

$$\Delta_4^{k_1}\Delta_6^{k_2}(D_{2,6}, f) = \sum_{s_1=0}^{k_2}\sum_{s_2=0}^{s_1}\sum_{t_1=0}^{k_1}\sum_{t_2=0}^{t_1}\binom{k_1}{s_1}\binom{s_1}{s_2}\binom{k_2}{t_1}\binom{t_1}{t_2}(t_2 + s_2)!(t_1 + s_1 - t_2 - s_2)!$$
$$\times (k_1 + k_2 - t_1 - s_1) \times O\left(\prod_{m=1}^{k_1+k_2}\frac{1}{f + 2m}\right).$$

Similarly to the proof of Lemma A.2.13 in Section A.2.2.12, by the Chu–Vandermonde identity, we obtain

$$\begin{aligned}\Delta_4^{k_1}\Delta_6^{k_2}(D_{2,6}, f) &= \sum_{s_1=0}^{k_2}\sum_{t_1=0}^{k_1}k_1!k_2!\binom{k_1 + k_2 - s_1 - t_1}{k_1 - s_1}\binom{s_1 + t_1}{s_1}(s_1 + t_1 + 1)\\ &= (k_1 + k_2 + 2)! \times O\left(\prod_{m=1}^{k_1+k_2}\frac{1}{f + 2m}\right),\end{aligned}$$

where we use $s_1 + s_2 + 1 \leq k_1 + k_2 + 1$ in the second equation.

A.2.2.15 Proof of Lemma A.2.16 (on Page 208)

We prove Lemma A.2.16 similarly to the proof of Lemma A.2.12 in Section A.2.2.11 by discussing $k_1 + k_2$ is odd and even, respectively.

(1) When $k_1 + k_2$ is odd, similarly to the analysis of (A.127), we assume without loss of generality that $k_2 \geq k_1$, and obtain

$$\begin{aligned} &\frac{(k_1+k_2+2)!}{k_1!k_2!} O\left(f^{-(k_1+k_2+1)/2} \prod_{t=1}^{k_1+k_2} \frac{1}{f+2t} \right) \\ \leq\ &\frac{2^{-(k_1+k_2-1)/2}}{k_1!k_2!} (k_1+k_2+2)(k_1+k_2+1) \prod_{t=1}^{(k_1+k_2+1)/2} t. \end{aligned} \tag{A.129}$$

Note that

$$\begin{aligned} &\frac{(k_1+k_2+2)(k_1+k_2+1)}{k_1!k_2!} \prod_{t=1}^{(k_1+k_2+1)/2} t \\ \leq\ &C \left(\frac{e^2}{k_1 k_2} \frac{k_1+k_2+3}{2e} \right)^{k_1} \left(\frac{e^2}{k_2^2} \frac{k_1+k_2+3}{2e} \right)^{(k_2-k_1)/2} (k_1+k_2+3)^{5/2} \\ \leq\ &C \left(\frac{2e}{k_1} \right)^{k_1} \left(\frac{2e}{k_2} \right)^{(k_2-k_1-5)/2} \left\{ \frac{2e(k_1+k_2+3)}{k_2} \right\}^{5/2}. \end{aligned} \tag{A.130}$$

When $k_2 - k_1 \geq 5$, we can see that (A.130) is bounded. When $k_2 - k_1 \leq 4$, we have

$$(\text{A.130}) \leq C \left(\frac{k_2}{k_1} \right)^{(5-k_2+k_1)/2} \left(\frac{2e}{k_1} \right)^{(k_1+k_2-5)/2} \left(\frac{k_1+k_2+3}{k_2} \right)^{5/2},$$

which suggests that (A.130) is bounded. In summary, we know (A.130) is bounded, and therefore (A.129) $= O\{2^{-(k_1+k_2-1)/2}\}$.

(2) When $k_1 + k_2$ is even, similar analysis can be applied, and then Lemma A.2.16 is proved.

A.2.2.16 Proof of Lemma A.1.5 (on Page 166)

We prove Lemma A.1.5 based on (A.57). In each testing problem, we have $|\tau_{1,k}+\upsilon_{1,k}|/|\eta\xi_{1,k}| = o(1)$; see Sections A.1.4.1–A.1.4.5. Then under the conditions of Lemma A.1.5, we can apply Lemma A.2.2 and obtain for $1 \leq k \leq K_1$,

$$
\begin{aligned}
&\log\Gamma\big\{\eta\xi_{1,k}(1-2it)+\tau_{1,k}+\upsilon_{1,k}\big\} \\
=&\left\{\eta\xi_{1,k}(1-2it)+\tau_{1,k}+\upsilon_{1,k}-\frac{1}{2}\right\}\log\left\{\eta\xi_{1,k}(1-2it)\right\}-\eta\xi_{1,k}(1-2it)+\log\sqrt{2\pi} \\
&+\sum_{l=1}^{L-1}\frac{(-1)^{l+1}B_{l+1}(\tau_{1,k}+\upsilon_{1,k})}{l(l+1)}\Big\{\eta\xi_{1,k}(1-2it)\Big\}^{-l}+O\Big(|\tau_{1,k}+\upsilon_{1,k}|^{L+1}/|\eta\xi_{1,k}|^{L}\Big).
\end{aligned}
$$

Applying similar expansion to $\log\Gamma(\eta\xi_{1,k}+\tau_{1,k}+\upsilon_{1,k})$, we obtain

$$
\begin{aligned}
&\log\Gamma\big\{\eta\xi_{1,k}(1-2it)+\tau_{1,k}+\upsilon_{1,k}\big\}-\log\Gamma\big(\eta\xi_{1,k}+\tau_{1,k}+\upsilon_{1,k}\big) \\
=&\left(\eta\xi_{1,k}+\tau_{1,k}+\upsilon_{1,k}-\frac{1}{2}\right)\log(1-2it)-2it\eta\xi_{1,k}\log\left\{\eta\xi_{1,k}(1-2it)\right\}+2it\eta\xi_{1,k} \\
&+\sum_{l=1}^{L-1}\frac{(-1)^{l+1}B_{l+1}(\tau_{1,k}+\upsilon_{1,k})}{l(l+1)(\eta\xi_{1,k})^{l}}\Big\{(1-2it)^{-l}-1\Big\}+O\Big(|\tau_{1,k}+\upsilon_{1,k}|^{L+1}/|\eta\xi_{1,k}|^{L}\Big).
\end{aligned}
$$

Similarly, for $1 \leq j \leq K_2$, we have

$$
\begin{aligned}
&\log\Gamma\big\{\eta\xi_{2,j}(1-2it)+\tau_{2,j}+\upsilon_{2,j}\big\}-\log\Gamma\big(\eta\xi_{2,j}+\tau_{2,j}+\upsilon_{2,j}\big) \\
=&\left(\eta\xi_{2,j}+\tau_{2,j}+\upsilon_{2,j}-\frac{1}{2}\right)\log(1-2it)-2it\eta\xi_{2,j}\log\left\{\eta\xi_{2,j}(1-2it)\right\}+2it\eta\xi_{2,j} \\
&+\sum_{l=1}^{L-1}\frac{(-1)^{l+1}B_{l+1}(\tau_{2,j}+\upsilon_{2,j})}{l(l+1)(\eta\xi_{2,j})^{l}}\Big\{(1-2it)^{-l}-1\Big\}+O\Big(|\tau_{2,j}+\upsilon_{2,j}|^{L+1}/|\eta\xi_{2,j}|^{L}\Big).
\end{aligned}
$$

Then by the form of $\varphi(t)$ in (A.57), we calculate

$$\begin{aligned}
\text{(A.57)} = {} & 2it\eta\left(\sum_{k=1}^{K_1}\xi_{1,k}\log\xi_{1,k} - \sum_{j=1}^{K_2}\xi_{2,j}\log\xi_{2,j}\right) \\
& + \left\{\sum_{k=1}^{K_1}(\xi_{1,k}+\tau_{1,k}+\upsilon_{1,k}-1/2) - \sum_{j=1}^{K_2}(\xi_{2,j}+\tau_{2,j}+\upsilon_{2,j}-1/2)\right\}\log(1-2it) \\
& - 2it\eta\left(\sum_{k=1}^{K_1}\xi_{1,k}\log\xi_{1,k} - \sum_{j=1}^{K_2}\xi_{2,j}\log\xi_{2,j}\right) - 2it\eta(\log\eta - 1)\left(\sum_{k=1}^{K_1}\xi_{1,k} - \sum_{j=1}^{K_2}\xi_{2,j}\right) \\
& + \sum_{l=1}^{L-1}\varsigma_l\left\{(1-2it)^{-l} - 1\right\} + O\left(\sum_{k=1}^{K_1}\frac{|\tau_{1,k}+\upsilon_{1,k}|^{L+1}}{|\eta\xi_{1,k}|^L} + \sum_{j=1}^{K_2}\frac{|\tau_{2,j}+\upsilon_{2,j}|^{L+1}}{|\eta\xi_{2,j}|^L}\right).
\end{aligned}$$

By the facts that $\tau_{1,k} = \eta\xi_{1,k}$, $\tau_{2,j} = \eta\xi_{2,j}$, and $\sum_{k=1}^{K_1}\xi_{1,k} = \sum_{j=1}^{K_2}\xi_{2,k}$, Lemma A.1.5 is proved.

A.2.3 Lemmas for Theorems 2.2.3 and 2.2.6

A.2.3.1 Proof of Lemma A.1.4 (on Page 149)

By (A.30) on Page 148,

$$\log\psi_1(s) = -\frac{pnti}{2}\log\frac{2e}{n} - \frac{pn(1-ti)}{2}\log(1-ti) + \log\frac{\Gamma_p\{(n-1)/2 - nti/2\}}{\Gamma_p\{(n-1)/2\}} + \mu_n ti,$$

where $t = s/(n\sigma_n)$. We next examine $\log\psi_1(s)$ by the following Lemma A.2.18.

Lemma A.2.18. *Let $\{p = p_n; n \geq 1\}$, $\{m = m_n; n \geq 1\}$, $\{t_n; n \geq 1\}$, and $\{s_n; n \geq 1\}$ satisfy that (i) $p_n \to \infty$ and $p_n = o(n)$; (ii) there exists $\epsilon \in (0,1)$ such that $\epsilon \leq m_n/n \leq \epsilon^{-1}$; (iii) $t = t_n = O(ns/p)$; (iv) $s = s_n = o(\min\{(n/p)^{1/2}, f^{1/6}\})$. Then as $n \to \infty$,*

$$\begin{aligned}
& \log\frac{\Gamma_p\left(\frac{m-1}{2}+ti\right)}{\Gamma_p\left(\frac{m-1}{2}\right)} \\
= {} & \beta_{m,1}ti - \beta_{m,2}t^2 + \beta_{m,3}(ti) + O\left(\frac{p^2t}{m^2}\right) + \left(\frac{1}{p}+\frac{p}{m}\right)O\left(\frac{p^2t^2}{m^2}\right) + O\left(\frac{p^2t^3}{m^3}\right),
\end{aligned}$$

where

$$\begin{aligned}
\beta_{m,1} =& -\left\{2p + \left(m - p - \frac{3}{2}\right)\log\left(1 - \frac{p}{m-1}\right)\right\};\\
\beta_{m,2} =& -\left\{\frac{p}{m-1} + \log\left(1 - \frac{p}{m-1}\right)\right\};\\
\beta_{m,3}(ti) =& p\left\{\left(\frac{m-1}{2} + ti\right)\log\left(\frac{m-1}{2} + ti\right) - \frac{m-1}{2}\log\frac{m-1}{2}\right\}.
\end{aligned}$$

Proof. Please see Section A.2.3.2 on Page 223. □

By (A.7) and $f = \Theta(p^2)$, we know $t = s/(n\sigma_n) = O(s/p)$. Thus we can apply Lemma A.2.18 and expand

$$\begin{aligned}
\log\frac{\Gamma_p\{(n-1)/2 - nti/2\}}{\Gamma_p\{(n-1)/2\}} =& -\frac{n\beta_{n,1}ti}{2} - \frac{\beta_{n,2}n^2t^2}{4} + \beta_{n,3}\left(-\frac{nti}{2}\right)\\
&+ O\left(\frac{p^2t}{n}\right) + \left(\frac{1}{p} + \frac{p}{n}\right)O\left(p^2t^2\right) + O\big(p^2t^3\big).
\end{aligned}$$

We next use the following Lemma A.2.19 to evaluate $\beta_{n,3}(-nti/2)$.

Lemma A.2.19. *When $p = p_n \to \infty$, $p = o(n)$, and $t = t_n = O(s/p)$ with $s = s_n = o(\min\{(n/p)^{1/2}, f^{1/6}\})$,*

$$\beta_{n,3}\bigg(-\frac{nti}{2}\bigg) = -\frac{pnti}{2}\log\frac{n}{2} + \frac{pn(1-ti)}{2}\log(1-ti) + \frac{pti}{2} + O\bigg(pt^2 + \frac{pt}{n}\bigg).$$

Proof. Please see Section A.2.3.3 on Page 225. □

It follows that

$$\begin{aligned}
\log\psi_1(s) =& -\frac{\{p(n-1) + n\beta_{n,1}\}ti}{2} - \frac{\beta_{n,2}n^2t^2}{4} + \mu_n ti\\
&+ O\bigg(\frac{p^2t}{n}\bigg) + \left(\frac{1}{p} + \frac{p}{n}\right)O\left(p^2t^2\right) + O\big(p^2t^3\big).
\end{aligned}$$

Since $\sigma_n^2 = \beta_{n,2}/2$, $\mu_n = \{p(n-1) + n\beta_{n,1}\}/2$, and $t = s/(n\sigma_n)$,

$$\log \psi_1(s) = -\frac{s^2}{2} + O\left(\frac{ps}{n}\right) + \left(\frac{1}{p} + \frac{p}{n}\right) O(s^2) + O\left(\frac{s^3}{p}\right),$$

where we use $t = O(s/p)$. As $\log \psi_0(s) = -s^2/2$, (A.31) is proved.

A.2.3.2 Proof of Lemma A.2.18 (on Page 221)

By the property of the multivariate gamma function; see, e.g., Theorem 2.1.12 in Muirhead (2009),

$$\log \frac{\Gamma_p\left(\frac{m-1}{2} + ti\right)}{\Gamma_p\left(\frac{m-1}{2}\right)} = \sum_{j=1}^{p} \log \frac{\Gamma\left(\frac{m-j}{2} + ti\right)}{\Gamma\left(\frac{m-j}{2}\right)}. \tag{A.131}$$

Then by Lemma A.2.3 on Page 184,

$$\begin{aligned}\log \frac{\Gamma\left(\frac{m-j}{2} + ti\right)}{\Gamma\left(\frac{m-j}{2}\right)} = \sum_{j=1}^{p} \Bigg[&\left(\frac{m-j}{2} + ti\right) \log\left(\frac{m-j}{2} + ti\right) - \left(\frac{m-j}{2}\right) \log\left(\frac{m-j}{2}\right) \\ &- ti - \frac{ti}{m-j} + O\left\{\frac{t+t^2}{(m-j)^2}\right\}\Bigg],\end{aligned} \tag{A.132}$$

as $m \to \infty$ uniformly for all $1 \le j \le p$. Note that $t/(m-j) = t/m + (t/m) \times \{j/(m-j)\}$, and then

$$\sum_{j=1}^{p} \frac{ti}{m-j} = \frac{pti}{m} + O\left(\frac{p^2}{m^2}\right) ti. \tag{A.133}$$

By (A.132) and (A.133), we obtain as $m \to \infty$,

$$\text{(A.131)} = \sum_{j=1}^{p}\left[\left(\frac{m-j}{2}+ti\right)\log\left(\frac{m-j}{2}+ti\right)-\left(\frac{m-j}{2}\right)\log\left(\frac{m-j}{2}\right)\right]$$

$$-\frac{(m+1)pti}{m}+O\left(\frac{p^2}{m^2}t+\frac{p}{m^2}t^2\right). \tag{A.134}$$

For $1 \le j \le p$, define

$$g_j(z) = \left(\frac{m-j}{2}+z\right)\log\left(\frac{m-j}{2}+z\right)-\left(\frac{m-1}{2}+z\right)\log\left(\frac{m-1}{2}+z\right),$$

where the real part of $z > -(m-p)/2$. It follows that the "$\sum_{j=1}^{p}$" term in the first row of (A.134) is equal to

$$p\left\{\left(\frac{m-1}{2}+ti\right)\log\left(\frac{m-1}{2}+ti\right)-\frac{m-1}{2}\log\frac{m-1}{2}\right\}+\sum_{j=1}^{p}\{g_j(ti)-g_j(0)\}. \tag{A.135}$$

To evaluate (A.135), we use the following Lemma A.2.20.

Lemma A.2.20. *Let $p = p_m$ such that $1 \le p < m$, $p \to \infty$ and $p/m \to 0$ as $m \to \infty$. When $t = t_m = O(ms/p)$ with $s = s_m = o(\min\{(m/p)^{1/2}, p^{1/3}\})$, we have that, as $m \to \infty$,*

$$\sum_{j=1}^{p}\{g_j(ti)-g_j(0)\} = \nu_{1,m}ti-\frac{\nu_{2,m}^2}{2}t^2+O\left(\frac{p^2t}{m^2}\right)+\left(\frac{1}{p}+\frac{p}{m}\right)O\left(\frac{p^2t^2}{m^2}\right)+O\Big(\frac{p^2t^3}{m^3}\Big),$$

where

$$\nu_{1,m} = \left(p-m+\frac{3}{2}\right)\log\left(1-\frac{p}{m-1}\right)-\frac{m-1}{m}p, \tag{A.136}$$

$$\nu_{2,m}^2 = -2\left\{\frac{p}{m-1}+\log\left(1-\frac{p}{m-1}\right)\right\}.$$

Proof. Please see Section A.2.3.4 on Page 226. ∎

Then by Lemma A.2.20,

$$
\begin{aligned}
\text{(A.135)} = p &\left\{ \left(\frac{m-1}{2} + ti\right) \log \left(\frac{m-1}{2} + ti\right) - \frac{m-1}{2} \log \frac{m-1}{2} \right\} \\
&+ \nu_{1,m} ti - \frac{\nu_{2,m}^2}{2} t^2 + O\left(\frac{p^2 t}{m^2}\right) + \left(\frac{1}{p} + \frac{p}{m}\right) O\left(\frac{p^2 t^2}{m^2}\right) + O\Big(\frac{p^2 t^3}{m^3}\Big).
\end{aligned}
$$

In summary, Lemma A.2.18 can be proved by noticing

$$
\begin{aligned}
&\beta_{m,1} = \nu_{1,m} - \frac{(m+1)p}{m}, \qquad \beta_{m,2} = \nu_{2,m}^2/2 \\
&\beta_{m,3}(ti) = p \left\{ \left(\frac{m-1}{2} + ti\right) \log \left(\frac{m-1}{2} + ti\right) - \frac{m-1}{2} \log \frac{m-1}{2} \right\}.
\end{aligned}
$$

A.2.3.3 Proof of Lemma A.2.19 (on Page 222)

By Taylor's series,

$$
\begin{aligned}
p^{-1}\beta_{n,3}(-nti/2) = &\; -\frac{nti}{2} \log \frac{n}{2} - \frac{nti}{2} \log \left(1 - ti - \frac{1}{n}\right) + \frac{n-1}{2} \log \left(1 - ti - \frac{ti}{n-1}\right) \\
= &\; -\frac{nti}{2} \log \frac{n}{2} - \frac{nti}{2} \log(1-ti) + \frac{nti}{2n(1-ti)} + O\Big(\frac{nt}{n^2}\Big) \\
&\; + \frac{n-1}{2} \log(1-ti) - \frac{n-1}{2} \frac{ti}{(n-1)(1-ti)} + \frac{n-1}{2} O\Big(\frac{t^2}{n^2}\Big) \\
= &\; -\frac{nti}{2} \log \frac{n}{2} + \frac{n(1-ti)-1}{2} \log(1-ti) + O\Big(\frac{t+t^2}{n}\Big).
\end{aligned}
$$

It follows that

$$
\beta_{n,3}(-nti/2) = -\frac{pnti}{2} \log \frac{n}{2} + \frac{pn(1-ti)}{2} \log(1-ti) + \frac{pti}{2} + O\Big(pt^2 + \frac{pt}{n}\Big).
$$

A.2.3.4 Proof of Lemma A.2.20 (on Page 224)

The first-order derivatives of $g_j(z)$ is

$$g_j^{(1)}(z) = \log\left(\frac{m-j}{2}+z\right) - \log\left(\frac{m-1}{2}+z\right),$$

and for $l \geq 2$, the l-th order derivatives of $g_j(z)$ is

$$\begin{aligned} g_j^{(l)}(z) &= (-1)^{l-2}(l-2)!\left\{\left(\frac{m-j}{2}+z\right)^{-(l-1)} - \left(\frac{m-1}{2}+z\right)^{-(l-1)}\right\} \\ &= (-1)^{l-2}(l-2)!\left(\frac{m-1}{2}+z\right)^{-(l-1)}\sum_{v=1}^{l-1}\binom{l-1}{v}\left(\frac{j-1}{m-j+2z}\right)^{v}. \end{aligned}$$

By Taylor's expansion, $g_j(ti) - g_j(0) = \sum_{l=1}^{\infty} g_j^{(l)}(0)z^l/l!$. In particular,

$$g_j^{(1)}(0) = \log(m-j) - \log(m-1), \qquad g_j^{(2)}(0) = \frac{2}{m-j} - \frac{2}{m-1}.$$

When $z = ti$, $t = t_m = O(ms/p)$, and $l \geq 3$, as $j - 1/(m-j+2z) = O(p/m) = o(1)$,

$$g_j^{(l)}(0)z^l/l! = O\left(\frac{1}{m^{l-1}}\frac{p}{m}t^l\right) = O\left(\frac{p}{m^l}\right)t^l.$$

As $t/m = O(s/p) = o(1)$,

$$\sum_{j=1}^{p}\{g_j(ti) - g_j(0)\} = \sum_{j=1}^{p} g_j^{(1)}(0)ti - \frac{1}{2}\sum_{j=1}^{p} g_j^{(2)}(0)t^2 + O\left(\frac{p^2t^3}{m^3}\right).$$

By Lemma A.2 in Jiang and Qi (2015),

$$\sum_{j=1}^{p} g_j^{(1)}(0) = \nu_{1,m} + O(\nu_{2,m}^2), \qquad \sum_{j=1}^{p} g_j^{(2)}(0) = \nu_{2,m}^2\left\{1 + O\left(\frac{1}{p}+\frac{p}{m}\right)\right\},$$

where $\nu_{1,m}$ and $\nu_{2,m}^2$ are defined in (A.136). In summary,

$$\sum_{j=1}^{p}\{g_j(ti)-g_j(0)\}=\nu_{1,m}ti-\frac{\nu_{2,m}^2}{2}t^2+O(\nu_{2,m}^2)t+\nu_{2,m}^2 O\left(\frac{1}{p}+\frac{p}{n}\right)t^2+O\Big(\frac{p^2t^3}{m^3}\Big).$$

Then Lemma A.2.20 follows by $\nu_{2,m}^2=O(p^2/m^2)$.

A.2.3.5 Proof of Lemma A.1.6 (on Page 182)

By Taylor's series, we have (A.70). In addition, for (A.71), note that we can write

$$p^{-1}\varrho_l(t)=\frac{l-1}{2}\log\left(1+\frac{lt}{l-1}\right)+\frac{lt}{2}\log\left(\frac{l-1}{2}+\frac{lt}{2}\right).$$

By Taylor's series $\log x=\log a+\sum_{l=1}^{L-1}(-1)^{l-1}l^{-1}(x/a-1)^l+O\{(x/a-1)^L\}$, we obtain

$$\begin{aligned}\frac{\varrho_l(t)}{p}&=\frac{l}{2}\log\left(1+t+\frac{t}{l-1}\right)-\frac{1}{2}\log\left(1+\frac{lt}{l-1}\right)+\frac{lt}{2}\log\left\{\frac{l(1+t)}{2}-\frac{1}{2}\right\}\\&=\frac{l}{2}\log(1+t)+\frac{lt}{2(l-1)(1+t)}-\frac{lt}{2(l-1)}+\frac{lt}{2}\log\frac{l(1+t)}{2}-\frac{t}{2(1+t)}+O\left(\frac{t}{l}+t^2\right)\\&=\frac{l(1+t)}{2}\log(1+t)+\frac{lt}{2}\log\frac{l}{2}-\frac{t}{2}+O\left(\frac{t}{l}+t^2\right).\end{aligned}$$

Then by $n=\sum_{j=1}^{k}n_j$, we have

$$-\varrho_n(t)+\sum_{j=1}^{k}\varrho_{n_j}(t)=\left(1-k-n\log n+\sum_{j=1}^{k}n_j\log n_j\right)\frac{tp}{2}+O\left(\frac{pt}{n}+pt^2\right).$$

A.3 Proofs of Theoretical Results in Section 2.3

A.3.1 Proof of Theorem 2.3.1

To derive the necessary and sufficient condition on the dimension of data, it is required to correctly understand the limiting behavior of the likelihood ratio test statistic under both low- and high-dimensional settings. In particular, we examine the limiting distribution of the likelihood ratio test statistic based on its moment generating function. For easy presentation in the technical proof, we let $n = N - 1$ in this section. Then we can write $T_0 = -n \log |\hat{R}_n|$. Under the conditions of Theorem 2.3.1, by Theorem 5.1.3 in Muirhead (2009) and Lemma 5.10 in Jiang and Yang (2013), we know that there exists a small constant $\delta_0 > 0$ such that for $h \in (-\delta_0, \delta_0)$,

$$\mathrm{E}\{\exp(h \times T_0)\} = \mathrm{E}\{|\hat{R}_n|^{-hn}\} = \left\{ \frac{\Gamma(n/2)}{\Gamma(n/2 - hn)} \right\}^p \times \frac{\Gamma_p(n/2 - hn)}{\Gamma_p(n/2)},$$

where $\Gamma(z)$ denotes the Gamma function, and $\Gamma_p(z)$ denotes the multivariate Gamma function satisfying $\Gamma_p(z) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma\{z - (j-1)/2\}$.

When p is fixed compared to N, by applying Stirling's approximation to the Gamma function, it can be shown that as $N \to \infty$, for any $h \in (-\delta_0, \delta_0)$, $\mathrm{E}\{\exp(h \times T_0)\}$ converges to $(1 - 2h)^{-f_0/2}$, which is the moment generating function of $\chi^2_{f_0}$; see, e.g., Bartlett (1950) and Section 5.1.2 of Muirhead (2009). It follows that $T_0 \xrightarrow{D} \chi^2_{f_0}$ by the continuity theorem. When $p \to \infty$, Jiang and Yang (2013) and Jiang and Qi (2015) derived an approximate expansion of the multivariate Gamma function $\Gamma_p(\cdot)$ when p increases with the sample size N, and then showed that for any $h \in (-\delta_0, \delta_0)$,

$$\mathrm{E}[\exp\{h(T_0 + n\mu_{n,0})/(n\sigma_{n,0})\}] \to \exp(h^2/2), \tag{A.137}$$

where δ_0 is a constant that is sufficiently small, $\exp(h^2/2)$ is the moment generating function of the standard normal random variable $\mathcal{N}(0, 1)$, and

$$\mu_{n,0} = (p - n + 1/2) \log\left(1 - \frac{p}{n}\right) - \frac{n-1}{n} p, \quad \sigma^2_{n,0} = -2\left\{\frac{p}{n} + \log\left(1 - \frac{p}{n}\right)\right\}.$$

Similarly to the proof in Section A.1.1.1, if the chi-squared approximation for T_0 holds, we know $\mathrm{E}[\exp\{h(T_0 - f_0)/\sqrt{2f_0}\}] \to \exp(h^2/2)$ for $h \in (-\delta_0, \delta_0)$, which, given

(A.137), is equivalent to $\sqrt{2f_0} \times (n\sigma_{n,0})^{-1} \to 1$ and $(f_0 + n\mu_{n,0}) \times (n\sigma_{n,0})^{-1} \to 0$. Therefore, similar to the analysis of (A.7) and (A.8) in Section A.1.1.1, we can obtain that the chi-squared approximation holds if and only if $p^2/n \to 0$. Moreover, similar to the analysis of (A.9) and (A.10) in Section A.1.1.1, we can obtain that the chi-squared approximation with the Bartlett correction holds if and only if $p^3/n^2 \to 0$. Recall that $N = n+1$ in this section. Thus, the same conclusions hold asymptotically by replacing n with N, that is, the chi-square approximations without and with the Bartlett correction hold if and only if $p^2/N \to 0$ and p^3/N^2, respectively.

A.3.2 Proof of Theorem 2.3.2

Similarly to the proof of Theorem 2.3.1, we next examine the limiting distribution of T' based on its moment generating function. In Theorem 2.3.2, testing $H'_{0,k} : \Sigma = \Lambda_k \Lambda_k^\top + \Psi_k$ when Λ_k and Ψ_k are given is equivalent to testing the null hypothesis $H_0 : \Sigma = \mathrm{I}_p$ by applying the data transformation $\Sigma_k^{-1/2} X_i$ with $\Sigma_k = \Lambda_k \Lambda_k^\top + \Psi_k$. Then by Corollary 8.4.8 in Muirhead (2009), under the null hypothesis, we have

$$\mathrm{E}\{\exp(h \times T')\} = \left(\frac{2e}{n}\right)^{-pnh} (1-2h)^{-pn(1-2h)/2} \times \frac{\Gamma_p\{n(1-2h)/2\}}{\Gamma_p(n/2)}, \quad \text{(A.138)}$$

where $n = N - 1$. When p is fixed compared to the sample size N, by applying Stirling's approximation to the Gamma function, it has been shown that as $N \to \infty$, (A.138) converges to $(1-2h)^{-f'/2}$, which is the moment generating function of $\chi^2_{f'}$ (Muirhead, 2009, Section 8.4.4), and therefore $T' \xrightarrow{D} \chi^2_{f'}$. When $p \to \infty$, by the proof of Lemma A.3.1 below, we have $\mathrm{E}[\exp\{h(T' + n\mu_n)/(n\sigma_n)\}] \to \exp(h^2/2)$, where

$$\mu_n = -p + (p - n + 1/2)\log\left(1 - \frac{p}{n}\right), \quad \sigma_n^2 = -2\left\{\frac{p}{n} + \log\left(1 - \frac{p}{n}\right)\right\}. \quad \text{(A.139)}$$

The conclusions can be obtained following similar analysis to Section A.3.1. To finish the proof, it remains to prove Lemma A.3.1 below.

Lemma A.3.1. *Under the conditions of Theorem 2.3.2, when $p \to \infty$ as $n = N-1 \to \infty$, we have $(T' + n\mu_n)/(n\sigma_n) \xrightarrow{D} \mathcal{N}(0,1)$ with μ_n and σ_n^2 in* (A.139).

Proof. It suffices to show that there exists a constant $\delta' > 0$ such that $\mathrm{E}[\exp\{h(T' + n\mu_n)/(n\sigma_n)\}] \to \exp(h^2/2)$ for all $|h| < \delta'$. Particularly, we let $s = h/(n\sigma_n)$, and prove $\log[\mathrm{E}\{\exp(sT')\}] \to h^2/2 - h\mu_n/\sigma_n$. By the moment generating function of T' in (A.138), we have

$$
\begin{aligned}
& \log\left[\mathrm{E}\{\exp(s \times T')\}\right] && \text{(A.140)}\\
= \quad & -pns\log(2e/n) - \frac{pn}{2}(1-2s)\log(1-2s) + \log\left\{\frac{\Gamma_p(n/2 - ns)}{\Gamma_p(n/2)}\right\}.
\end{aligned}
$$

We next derive the approximate expansion of (A.140) by discussing two cases.

Case 1: $\lim p/n \to C \in (0,1]$. Under this case, we utilize the approximate expansion of multivariate gamma function in Lemma 5.4 of Jiang and Yang (2013). To apply the result, we first show that the conditions are satisfied. Specifically, define $r_n^2 = -\log(1 - p/n)$, and we have

$$
(-ns)^2 \times r_n^2 = -\frac{h^2}{\sigma_n^2}\log(1 - p/n) \to \begin{cases} \dfrac{h^2}{2} \times \dfrac{\log(1-C)}{C + \log(1-C)}, & \text{if } C \in (0,1); \\ \dfrac{h^2}{2}, & \text{if } C = 0. \end{cases}
$$

Therefore, $-ns = O(1/r_n)$, and then Lemma 5.4 in Jiang and Yang (2013) can be applied to expand (A.140). It follows that

$$
\begin{aligned}
\text{(A.140)} \quad = \quad & -pns\log(2e/n) - \frac{pn}{2}(1-2s)\log(1-2s) \\
& -pns\log\{n/(2e)\} + r_n^2\left\{(-ns)^2 - (p - n + 1/2)(-ns)\right\} + o(1).
\end{aligned}
$$

By Taylor's expansion $(1-2s)\log(1-2s) = -2s + 2s^2 + O(s^3)$ for $s \in (0,1)$, we obtain

$$
\begin{aligned}
\text{(A.140)} \quad = \quad & -\frac{pn}{2}\left\{-2s+2s^2+O(s^3)\right\} \\
& -\log\left(1-\frac{p}{n}\right)\left\{n^2s^2+(p-n+1/2)ns\right\}+o(1) \\
= \quad & s^2\left\{-pn-n^2\log\left(1-\frac{p}{n}\right)\right\}+s\left\{pn-(p-n+1/2)\log\left(1-\frac{p}{n}\right)\right\}+o(1).
\end{aligned}
$$

With $s=h/(n\sigma_n)$, we have $\log(\mathrm{E}[\exp\{hT'/(n\sigma_n)\}])=h^2/2-h\mu_n/\sigma_n+o(1)$.

Case 2: $\lim p/n=0$. Under this case, we utilize the approximate expansion of multivariate gamma function in Proposition A.1 of Jiang and Qi (2015). To apply the result, we first show that the conditions are satisfied. Particularly, as $\sigma_n^2=p^2n^{-2}\{1+o(1)\}$, we have $-ns\times p/n=-ph(n\sigma_n)^{-1}=h\{1+o(1)\}$. Therefore, $-ns=O(n/p)$, and we can apply Proposition A.1 in Jiang and Qi (2015) to expand (A.140). It follows that

$$
\log\left\{\frac{\Gamma_p(n/2-ns)}{\Gamma_p(n/2)}\right\}=\gamma_{n,1}(-ns)+\gamma_{n,2}(-ns)^2+\gamma_{n,3}+o(1),
$$

where

$$
\begin{aligned}
\gamma_{n,1} &= -\left\{2p+(n-p-1/2)\log\left(1-p/n\right)\right\}, \\
\gamma_{n,2} &= -\left\{p/n+\log\left(1-p/n\right)\right\}, \\
\gamma_{n,3} &= p\left\{(n/2-ns)\log\left(n/2-ns\right)-(n/2)\log\left(n/2\right)\right\}.
\end{aligned}
$$

Note that $\gamma_{n,3}=(pn/2)(1-2s)\log(1-2s)-pns\log(n/2)$. Then we have

$$
\begin{aligned}
\text{(A.140)} = \; & -pns\log\left(\frac{2e}{n}\right)-\frac{pn}{2}(1-2s)\log(1-2s)-\gamma_{n,1}ns+\gamma_{n,2}n^2s^2+\gamma_{n,3}+o(1) \\
= \; & -(p+\gamma_{n,1})ns+\gamma_{n,2}n^2s^2+o(1),
\end{aligned}
$$

which gives $\log(\mathrm{E}[\exp\{hT'/(n\sigma_n)\}]) = h^2/2 + \mu_n h/\sigma_n + o(1)$ by $s = h/(n\sigma_n)$.

Finally, for a general sequence $\{p/n\}$, to prove that $(T' + n\mu_n)/(n\sigma_n)$ converges in distribution to $\mathcal{N}(0, 1)$, it suffices to show that every subsequence has a further subsequence that converges in distribution to $\mathcal{N}(0, 1)$. By the boundedness of p/n and the Bolzano-Weierstrass theorem, we can further take a subsequence such that p/n has a limit and the arguments above can be applied. In summary, Lemma A.3.1 is proved. $\square$

A.4 Proofs of Theoretical Results in Section 2.4

A.4.1 Proof of Theorem 2.4.1

Proof of Part (i): $\boldsymbol{mr \to \infty}$ We prove the conclusion for $mr \to \infty$ in Theorem 2.4.1 based on Theorem 2.4.3. When (p, m, r) are all fixed, we know that $-2 \log L_n \xrightarrow{D} \chi^2_{mr}$ as $n \to \infty$. Note that $\mathrm{E}(\chi^2_{mr}) = mr$, $\mathrm{var}(\chi^2_{mr}) = 2mr$, and when $mr \to \infty$, $(\chi^2_{mr} - mr)/\sqrt{2mr} \xrightarrow{D} \mathcal{N}(0, 1)$. It follows that $P(\chi^2_{mr} > \sqrt{2mr} z_\alpha + mr) \to \alpha$ and

$$\chi^2_{mr}(\alpha) = \sqrt{2mr} \times \{z_\alpha + o(1)\} + mr, \tag{A.141}$$

where z_α denotes the upper α-quantile of $\mathcal{N}(0, 1)$.

We define the asymptotic regime $\mathcal{R}_A = \{(p, m, r, n) : n > p + m,\ p \geq r,\ mr \to \infty, \text{ and } \max\{p, m, r\}/n \to 0 \text{ as } n \to \infty\}$. Under the asymptotic regime $\mathcal{R}_A$, Theorem 2.4.3 shows that $(-2 \log L_n + \mu_n)/(n\sigma_n) \xrightarrow{D} \mathcal{N}(0, 1)$. Note that

$$P\{-2 \log L_n > \chi^2_{mr}(\alpha)\} = P\Big\{\frac{-2 \log L_n + \mu_n}{n\sigma_n} > \frac{\chi^2_{mr}(\alpha) + \mu_n}{n\sigma_n}\Big\}.$$

Thus when $n \to \infty$, $P\{-2 \log L_n > \chi^2_{mr}(\alpha)\} \to \alpha$ is equivalent to

$$\frac{\chi^2_{mr}(\alpha) + \mu_n}{n\sigma_n} \to z_\alpha, \quad \text{as } n \to \infty. \tag{A.142}$$

When $mr \to \infty$, by (A.141), we know (A.142) is equivalent to

$$\frac{\sqrt{2mr} \times \{z_\alpha + o(1)\} + mr + \mu_n}{n\sigma_n} \to z_\alpha, \quad \text{as } n \to \infty. \tag{A.143}$$

(A.143) holds for any significance level α if and only if $n\sigma_n = \sqrt{2mr}\{1 + o(1)\}$ and $(\mu_n + mr)/\sqrt{2mr} = o(1)$.

Next we will prove that under $\mathcal{R}_A$, $n\sigma_n = \sqrt{2mr}\{1+o(1)\}$ in the first step, derive the form of μ_n in the second step, and obtain the conclusion in the third step.

Step 1. Note that

$$\sigma_n^2 = 2\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)}.$$

By the Taylor expansion, $\log(1-a) = -a - a^2/2 - a^3/3 + O(a^4)$ for $a = o(1)$. Under $\mathcal{R}_A$, we know that $p/n, m/n, r/n \to 0$ and $r/(n-p-m) \to 0$. Then we have

$$\begin{aligned} &\log\frac{n+r-p-m}{n-p-m} \\ =\ & \frac{r}{n-p-m} - \frac{1}{2}\frac{r^2}{(n-p-m)^2} + \frac{1}{3}\frac{r^3}{(n-p-m)^3} + O\Big(\frac{r^4}{n^4}\Big), \end{aligned} \tag{A.144}$$

and similarly,

$$-\log\frac{n-p+r}{n-p} \quad = \quad -\frac{r}{n-p} + \frac{1}{2}\frac{r^2}{(n-p)^2} - \frac{1}{3}\frac{r^3}{(n-p)^3} + O\Big(\frac{r^4}{n^4}\Big). \tag{A.145}$$

Then

$$\begin{aligned} &(\text{A.144}) + (\text{A.145}) \\ =\ & \frac{rm}{(n-p-m)(n-p)} - \frac{1}{2}r^2 \times \frac{m(2n-2p-m)}{(n-p-m)^2(n-p)^2} + O\Big\{\frac{r^3(m+r)}{n^4}\Big\}. \end{aligned} \tag{A.146}$$

We next examine the first two terms in (A.146). Note that for $a = o(1)$ and

$b = o(1)$, $1/(1-a) = 1 + a + O(a^2)$ and $1/\{(1-a)(1-b)\} = 1 + a + b + O(a^2 + b^2)$. Then for the first term in (A.146), we have

$$\frac{rm}{(n-p-m)(n-p)} = \frac{rm}{n^2}\Big\{1 + \frac{2p+m}{n} + O\Big(\frac{p^2+m^2}{n^2}\Big)\Big\}. \tag{A.147}$$

In addition, note that for $a = o(1)$ and $b = o(1)$, $1/\{(1-a)^2(1-b)^2\} = 1 + 2a + 2b + O(a^2 + b^2)$. Then for the second term in (A.146), we have

$$-\frac{1}{2}r^2 \times \frac{m(2n-2p-m)}{(n-p-m)^2(n-p)^2} = -\frac{mr^2}{n^3}\Big\{1 + \frac{3p+3m/2}{n} + O\Big(\frac{p^2+m^2}{n^2}\Big)\Big\}. \tag{A.148}$$

Combining (A.147) and (A.148), we obtain

$$\text{(A.146)} = \frac{rm}{n^2} + \frac{rm}{n^2}\Big\{\frac{2p+m-r}{n}\Big\} + O\Big\{\frac{mr(m^2+r^2+p^2)}{n^4}\Big\}. \tag{A.149}$$

We then know that $\sigma_n^2 = 2 \times \text{(A.146)} = (2mr/n^2) \times \{1 + o(1)\}$, and thus $n\sigma_n = \sqrt{2mr}\{1 + o(1)\}$.

Step 2. In this step, we derive the asymptotic form of μ_n under the asymptotic region $\mathcal{R}_A$. Particularly, note that

$$\begin{aligned} \mu_n = \; & n(n-m-p+r-1/2)\log\Big(1 - \frac{m}{n+r-p}\Big) && \text{(A.150)} \\ & + n(n-p-1/2)\log\Big(1 - \frac{r}{n+r-p}\Big) && \text{(A.151)} \\ & - n(n-m-p-1/2)\log\Big(1 - \frac{m+r}{n+r-p}\Big). && \text{(A.152)} \end{aligned}$$

By the Taylor expansion,

$$\text{(A.150)}/n = -\sum_{k=1}^{\infty}\frac{1}{k}\frac{m^k}{(n+r-p)^{k-1}} + \sum_{k=1}^{\infty}\frac{1}{k}\frac{m^{k+1}}{(n+r-p)^k} + \sum_{k=1}^{\infty}\frac{1}{k}\frac{m^k/2}{(n+r-p)^k}.$$

Similarly, by applying the Taylor expansion to (A.151) and (A.152), we have

$$
\begin{aligned}
&(\text{A.150}) + (\text{A.151}) + (\text{A.152})\}/n \\
=&\sum_{k=1}^{\infty}\frac{1}{k}\frac{(m+r)^k - m^k - r^k}{(n+r-p)^{k-1}} - \sum_{k=1}^{\infty}\frac{1}{k}\frac{(m+r)^{k+1} - m^{k+1} - r^{k+1}}{(n+r-p)^k}, && (\text{A.153}) \\
&-\frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{k}\frac{(m+r)^k - m^k - r^k}{(n+r-p)^k}. && (\text{A.154})
\end{aligned}
$$

Note that

$$
\begin{aligned}
(\text{A.153}) = &-\frac{mr}{(n+r-p)} - \frac{mr(m+r)}{2(n+r-p)^2} - \frac{mr(m^2/3 + mr/2 + r^2/3)}{(n+r-p)^3} && (\text{A.155}) \\
&-\sum_{k=4}^{\infty}\frac{1}{k(k+1)}\frac{(m+r)^{k+1} - m^{k+1} - r^{k+1}}{(n+r-p)^k}, && (\text{A.156})
\end{aligned}
$$

where

$$
\begin{aligned}
|(\text{A.156})| &= \sum_{k=4}^{\infty}\frac{1}{k(k+1)}\frac{\sum_{q=1}^{k}\binom{k+1}{q}m^q r^{k+1-q}}{(n+r-p)^k} \\
&\le \frac{mr}{n+r-p}\sum_{k=4}^{\infty}\frac{1}{k}\times\frac{2^{k+1}(\max\{m,r\})^{k-1}}{(n+r-p)^{k-1}} \\
&= \frac{mr}{n+r-p}O\Big\{\Big(\frac{\max\{m,r\}}{n+r-p}\Big)^3\Big\} = O(1)\frac{mr(m^3+r^3)}{n^4},
\end{aligned}
$$

where in the last two equations, we use the property of Taylor expansion and the condition that $\max\{p,m,r\} = o(n)$. Therefore, $n\times(\text{A.156}) = mr\times O\{(m^3+r^3)/n^3\}$. Similarly, we also obtain $(\text{A.154}) = O(mr/n^2)$. In summary,

$$
\begin{aligned}
\mu_n &= (\text{A.150}) + (\text{A.151}) + (\text{A.152}) = n\times\{(\text{A.153}) + (\text{A.154})\} \\
&= -\frac{nmr}{n+r-p} - \frac{1}{2}\frac{nmr(m+r)}{(n+r-p)^2} - \frac{nmr(m^2/3 + mr/2 + r^2/3)}{(n+r-p)^3} \\
&\quad + O(1)\frac{mr(m^3+r^3)}{n^3} + O(mr/n) \\
&= -mr\Big\{1 + \frac{p-r}{n}\Big\} - \frac{mr(m+r)}{2n} + o(1)mr\times\frac{p+m+r}{n}.
\end{aligned}
$$

Step 3. Combining the results in *Step 1.* and *Step 2.*, we have

$$\frac{\mu_n + mr}{\sqrt{2mr}} = -\frac{\sqrt{mr}}{\sqrt{2}}\Big(\frac{p + m/2 - r/2}{n}\Big) \times \{1 + o(1)\},$$

which converges to 0, if and only if $\lim_{n\to\infty} \sqrt{mr}(p + m/2 - r/2)n^{-1} = 0$.

Proof of Part (ii): *mr* is finite By Muirhead (2009), $\phi_1(t) = \mathrm{E}\{\exp(-2it \log L_n)\}$, the characteristic function of $-2 \log L_n$, satisfies

$$\log \phi_1(t) = -\frac{mr}{2} \log(1 - 2it) + \sum_{l=1}^{\infty} \varsigma_l\{(1 - 2it)^{-l} - 1\}, \tag{A.157}$$

where

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)}\Big[\sum_{k=1}^{m}\Big\{\frac{\mathbb{B}_{l+1}\{(1-k-p)/2\}}{(n/2)^l} - \frac{\mathbb{B}_{l+1}\{(1-k+r-p)/2\}}{(n/2)^l}\Big\}\Big],$$

and $\mathbb{B}_{l+1}(\cdot)$ is the Bernoulli polynomials which takes the form $\mathbb{B}_{l+1}(z) = \sum_{v=0}^{l+1} c_v z^v$. We next estimate the order of ς_l with respect to n. Note that for any z_1 and z_2,

$$\mathbb{B}_{l+1}(z_1) - \mathbb{B}_{l+1}(z_2) = (z_1 - z_2)\sum_{v=1}^{l+1}\sum_{w=1}^{v} c_v \binom{v}{w}(z_1 - z_2)^{w-1} z_2^{v-w}. \tag{A.158}$$

Let $z_1 = (1 - k - p)/2$ and $z_2 = (1 - k + r - p)/2$. Then we have $z_1 - z_2 = (-r)/2$. When m and r are finite, the order of ς_l with respect to n is $O\{(p/n)^l\}$. When $p/n \to 0$, by the expansion (A.157), we have $\phi_1(t) = (1 - 2it)^{-mr/2}\{1 + o(1)\}$. Then $-2 \log L_n \xrightarrow{D} \chi^2_{mr}$ as $n \to \infty$. When p/n is bounded from 0 below, (A.157) does not converge to $-2^{-1} mr \log(1 - 2it)$ generally for all t. Then the approximation $-2 \log L_n \xrightarrow{D} \chi^2_{mr}$ fails.

A.4.2 Proof of Theorem 2.4.2

Part (i): $mr \to \infty$ When (p, m, r) are all fixed, we know that with the Bartlett correction factor $\rho = 1-(p-r/2+m/2+1/2)/n$, $-2\rho\log L_n \xrightarrow{D} \chi^2_{mr}$ as $n \to \infty$. Note that under $\mathcal{R}_A = \{(p,m,r,n) : n > p+m,\ p \geq r,\ mr \to \infty,\ \text{and}\ \max\{p,m,r\}/n \to 0\ \text{as}\ n \to \infty\}$, $\rho = 1+o(1)$. Then similarly to the proof of Theorem 2.4.1 in Section A.4.1, we know that under $\mathcal{R}_A$, $P\{-2\rho\log L_n > \chi^2_{mr}(\alpha)\} \to \alpha$ holds for any given significance level α if and only if $n\sigma_n = \sqrt{2mr}\{1+o(1)\}$ and $(\mu_n + mr/\rho)/\sqrt{2mr} = o(1)$.

Following the argument in Section A.4.1, we know that under $\mathcal{R}_A$, $n\sigma_n = \sqrt{2mr}\{1+o(1)\}$. In addition, by the Taylor expansion,

$$\begin{aligned}\frac{mr}{\rho} =& \frac{mr}{1-(p+m/2-r/2+1/2)/n}\\ =& \frac{nmr}{n-p+r} + \frac{nmr(m+r)}{2(n+r-p)^2} + \frac{nmr(m+r)^2}{4(n-p+r)^3} + O\Big(\frac{mr(m^3+r^3)}{n^3}\Big) + O\Big(\frac{mr}{n}\Big).\end{aligned}$$

It follows that under $\mathcal{R}_A$, by (A.150)–(A.156),

$$\frac{\mu_n - (-mr/\rho)}{\sqrt{2mr}} = -\frac{\sqrt{mr}(m^2+r^2)}{12\sqrt{2}n^2} + o(1) \times \frac{\sqrt{mr}(m^2+r^2)}{n^2} + O(\sqrt{mr}/n), \quad \text{(A.159)}$$

We thus know that (A.159) $\to 0$ if and only if $\sqrt{mr}(m^2+r^2)/n^2 \to 0$.

Part (ii): mr is finite By Muirhead (2009), for the LRT with Bartlett correction, the characteristic function of $-2\rho\log L_n$ is $\phi_2(t) = \mathrm{E}\{\exp(-2it\rho\log L_n)\}$. Moreover, we have $\log\phi_2(t) = -\frac{mr}{2}\log(1-2it) + \sum_{l=1}^{\infty}\tilde{\varsigma}_l\{(1-2it)^{-l}-1\}$,, where

$$\tilde{\varsigma}_l = \frac{(-1)^{l+1}}{l(l+1)}\Big[\sum_{k=1}^{m}\Big\{\frac{\mathbb{B}_{l+1}(\tilde{z}_{k,1})}{(\rho n/2)^l} - \frac{\mathbb{B}_{l+1}(\tilde{z}_{k,2})}{(\rho n/2)^l}\Big\}\Big],$$

$\tilde{z}_{k,1} = (1-\rho)n/2 + (1-k-p)/2$ and $\tilde{z}_{k,2} = (1-\rho)n/2 + (1-k+r-p)/2$. Since $\rho = 1-(p-r/2+m/2+1/2)/n$,

$$
\begin{aligned}
\tilde{z}_{k,1} =& (p-r/2+m/2+1/2)/2 + (1-k-p)/2 = (3-r+m)/4,\\
\tilde{z}_{k,2} =& (p-r/2+m/2+1/2)/2 + (1-k+r-p)/2 = (3+r+m)/4.
\end{aligned}
$$

In addition, $\rho n = n-(p-r/2+m/2+1/2)$. Therefore, by the expansion in (A.158), when m and r are fixed and $n-p\to\infty$, we have $\log\phi_2(t) = -2^{-1}mr\log(1-2it) + O\{(n-p)^{-1}\}$ and $\phi_2(t) = (1-2it)^{-mr/2}[1+O\{(n-p)^{-1}\}]$. It follows that when m and r are fixed and $n-p\to\infty$, $-2\rho\log L_n \xrightarrow{D} \chi^2_{mr}$. On the other hand, when $n-p$ is fixed, by the expansion in (A.158), we know $\tilde{\varsigma}_l$ is of constant order in n, and thus $\sum_{l=1}^{\infty}\tilde{\varsigma}_l\{(1-2it)^{-l}-1\}$ is not ignorable generally for all t. We then know the approximation $-2\rho\log L_n \xrightarrow{D} \chi^2_{mr}$ fails.

A.4.3 Proof of Theorem 2.4.3

It is sufficient to show

$$
E\exp\left\{\frac{\log L_n - \mu_n/2}{n\sigma_n/2}s\right\} \to \exp\{s^2/2\}, \tag{A.160}
$$

as $n\to\infty$ and $|s|<1$, where σ_n^2 and μ_n are defined in Theorem 2.4.3. Equivalently, it suffices to show that for any subsequence $\{n_k\}$, there is a further subsequence $\{n_{k_j}\}$ such that $H_{n_{k_j}}$ converges to $\mathcal{N}(0,1)$ in distribution as $j\to\infty$. In the following, the further subsequence is selected in a way such that the subsequential limits of some bounded quantities (to be specified in the proof below) exist, which is guaranteed by Bolzano-Weierstrass Theorem. Therefore, we only need to verify the theorems by assuming that the limits for these bounded quantities exist. In the following, we give the proof by discussing two settings $r\geq m$ and $m\geq r$ separately.

Case 1. When $r \geq m$ and $r \to \infty$. By Lemma A.4.2, under the null hypothesis, the distribution of L_n can be reexpressed as the distribution of a product of independent beta random variables. Let $h = 2s/(n\sigma_n)$, by Lemma A.4.1, then under the null hypothesis, L_n's hth moment can be written as

$$E \exp\left\{\frac{\log L_n}{n\sigma_n/2} s\right\} = E(L_n^h) = \frac{\Gamma_m\{\frac{1}{2}n(1+h) - \frac{1}{2}p\}\Gamma_m\{\frac{1}{2}(n+r-p)\}}{\Gamma_m\{\frac{1}{2}(n-p)\}\Gamma_m\{\frac{1}{2}n(1+h) + \frac{1}{2}(r-p)\}}, \quad \text{(A.161)}$$

where $\Gamma_m(a)$, $a \in \mathbb{C}$ and $\text{Re}(a) > (m-1)/2$, is the multivariate Gamma function defined to be

$$\Gamma_m(a) = \int_{A>0} e^{-\text{tr}(A)} \det A^{a-(m+1)/2} (dA). \quad \text{(A.162)}$$

The above integration is taken over the space of positive definite $m \times m$ matrices, i.e., $\{A_{m\times m} : A \succ 0\}$; and $\text{tr}(A)$ is the trace of A. Note that when $m = 1$, $\Gamma_m(a)$ becomes the usual definition of Gamma function. By Lemma A.4.3, $\Gamma_m(a)$ can be written as a product of ordinary Gamma functions as

$$\Gamma_m(a) = \pi^{m(m-1)/4} \prod_{j=1}^{m} \Gamma\{a - (j-1)/2\}.$$

Note that $n > m+p$ and $r \geq 1$. Thus the limits of $m/(n+r-p)$ and $m/(n-p)$ are in $[0,1]$ for all n. Applying the subsequence argument above, for any subsequence $\{n_k\}$, we take a further subsequence n_{k_j} such that $m_{k_j}/(n_{k_j} + r_{k_j} - p_{k_j})$ and $m_{k_j}/(n_{k_j} - p_{k_j})$ converge to some constants in $[0,1]$. Thus without loss of generality, we consider the cases when $m/(n+r-p)$ and $m/(n-p)$ converge to some constants in $[0,1]$. Next we give the proof by discussing different cases below.

Case 1.1 If $m/(n+r-p) \to \gamma > 0$, this implies that $m \to \infty$ as $n \to \infty$. And as $r \geq m$ and $n > p+m$, we know $m/(n+r-p) \leq 1/2$, then $\gamma \in (0, 1/2]$. Since

$1 \geq m/(n-p) \geq m/(n+r-p)$, then $m/(n-p) \to \gamma' \in (0,1]$.

If $\gamma' \in (0,1)$, $nh \times [-\log\{1-m/(n-p)\}]^{1/2} = O(1)$, which satisfies the assumption of Lemma 5.4 in Jiang and Yang (2013). If $\gamma' = 1$, as

$$\sigma_n^2 = 2\log\Big(1 - \frac{m}{n+r-p}\Big) - 2\log\Big(1 - \frac{m}{n-p}\Big), \tag{A.163}$$

and $m/(n+r-p) \to \gamma \in (0, 1/2]$, we know σ_n^2 has leading order $\log\{1 - m/(n-p)\}$. Then as $nh\sigma_n = O(1)$ by definition, we also know $nh \times [-\log\{1 - m/(n-p)\}]^{1/2} = O(1)$, which satisfies the assumption of Lemma 5.4 in Jiang and Yang (2013). Following the lemma, we have

$$\begin{aligned}\log \frac{\Gamma_m\{\frac{1}{2}n(1+h) - \frac{1}{2}p\}}{\Gamma_m\{\frac{1}{2}(n-p)\}} =& \log \frac{\Gamma_m\{\frac{1}{2}(n-p) + \frac{1}{2}nh\}}{\Gamma_m\{\frac{1}{2}(n-p)\}} \\ =& -\Big\{\frac{n^2h^2}{4} + \frac{nh}{2}\Big(n - m - p - \frac{1}{2}\Big)\Big\}\log\Big(1 - \frac{m}{n-p}\Big) \\ &+ \frac{mnh}{2}\{\log(n-p) - \log 2e\} + o(1),\end{aligned} \tag{A.164}$$

and similarly, we can obtain

$$\begin{aligned}\log \frac{\Gamma_m\{\frac{1}{2}(n+r-p)\}}{\Gamma_m\{\frac{1}{2}n(1+h) + \frac{1}{2}(r-p)\}} =& \log \frac{\Gamma_m\{\frac{1}{2}(n+r-p)\}}{\Gamma_m\{\frac{1}{2}(n+r-p) + \frac{1}{2}nh\}} \\ =& \Big\{\frac{n^2h^2}{4} + \frac{nh}{2}\Big(n + r - m - p - \frac{1}{2}\Big)\Big\}\log\Big(1 - \frac{m}{n+r-p}\Big) \\ &- \frac{mnh}{2}\{\log(n+r-p) - \log 2e\} + o(1).\end{aligned} \tag{A.165}$$

Combining (A.161), (A.164) and (A.165), we have

$$\begin{aligned}\log E\exp\Big\{\frac{\log L_n}{n\sigma_n/2}s\Big\} =& \frac{n^2h^2}{4}\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)} + \frac{h\mu_n}{2} + o(1) \\ =& \frac{s^2}{2} + \frac{h\mu_n}{2} + o(1),\end{aligned}$$

where

$$\mu_n = n(n-m-p-1/2)\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)} + nr\log\frac{(n+r-p-m)}{(n+r-p)} + nm\log\frac{(n-p)}{(n+r-p)}.$$

Therefore, $\log E\exp\left\{\frac{\log L_n-\mu_n/2}{n\sigma_n/2}s\right\} = s^2/2+o(1)$ is proved.

Case 1.2 We discuss the case when $m/(n+r-p)\to 0$ and $m/(n-p)\to 0$ below. By Lemma A.4.6, we know that when $n-p\to\infty$ and $r\to\infty$,

$$\begin{aligned}
&\log\frac{\Gamma_m\{\frac{1}{2}n(1+h)-\frac{1}{2}p\}}{\Gamma_m\{\frac{1}{2}(n-p)\}} \\
&= -\left\{2m+\left(n-p-m-\frac{1}{2}\right)\log\left(1-\frac{m}{n-p}\right)\right\}\frac{nh}{2} \\
&\quad -\left\{\frac{m}{n-p}+\log\left(1-\frac{m}{n-p}\right)\right\}\frac{n^2h^2}{4} \\
&\quad + m\left\{\frac{(n-p+nh)}{2}\log\frac{(n-p+nh)}{2}-\frac{(n-p)}{2}\log\frac{(n-p)}{2}\right\}+o(1),
\end{aligned} \tag{A.166}$$

and

$$\begin{aligned}
&\log\frac{\Gamma_m\{\frac{1}{2}(n+r-p)\}}{\Gamma_m\{\frac{1}{2}n(1+h)+\frac{1}{2}(r-p)\}} \\
&= \left\{2m+\left(n+r-p-m-\frac{1}{2}\right)\log\left(1-\frac{m}{n+r-p}\right)\right\}\frac{nh}{2} \\
&\quad - m\left\{\frac{(n+r-p+nh)}{2}\log\frac{(n+r-p+nh)}{2}-\frac{(n+r-p)}{2}\log\frac{(n+r-p)}{2}\right\} \\
&\quad + \left\{\frac{m}{n+r-p}+\log\left(1-\frac{m}{n+r-p}\right)\right\}\frac{n^2h^2}{4}+o(1).
\end{aligned} \tag{A.167}$$

By Taylor expansion of the log function, we have

$$\begin{aligned}
\sigma_n^2 &= 2\log\left(1-\frac{m}{n+r-p}\right)-2\log\left(1-\frac{m}{n-p}\right) \\
&= \frac{2mr}{(n-p)(n+r-p)}\{1+o(1)\},
\end{aligned} \tag{A.168}$$

where the second order terms of Taylor expansion of the log functions is ignorable as $m = o(n-p)$. Also, as $r \to \infty$,

$$h = \frac{s}{n\sigma_n/2} = \frac{s\sqrt{2(n-p)(n+r-p)}}{n\sqrt{mr}}\{1+o(1)\} \to 0. \tag{A.169}$$

Therefore, combining (A.161), (A.166) and (A.167), we obtain

$$\begin{aligned}
&\log E \exp\left\{\frac{\log L_n}{n\sigma_n/2}s\right\} \qquad\qquad (\text{A.170})\\
= \;& \frac{n^2h^2}{4}\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)} + \frac{n^2h^2}{4}\left(\frac{m}{n+r-p} - \frac{m}{n-p}\right)\\
&+\frac{nh}{2}(n-m-p-1/2)\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)}\\
&+\frac{nh}{2}r\log\frac{(n+r-p-m)}{(n+r-p)} + \frac{nh}{2}m\log\frac{n-p+nh}{n+r-p+nh}\\
&+\frac{m(n+r-p)}{2}\log\frac{n+r-p}{n+r-p+nh}\\
&+\frac{m(n-p)}{2}\log\frac{n-p+nh}{n-p} + o(1).
\end{aligned}$$

We then analyze the terms in (A.170) separately. By (A.169),

$$\begin{aligned}
&\frac{n^2h^2}{4}\left(\frac{m}{n+r-p} - \frac{m}{n-p}\right)\\
= \;& -\frac{s^2(n-p)(n+r-p)}{2mr} \times \frac{mr}{(n-p)(n+r-p)}\{1+o(1)\}\\
= \;& -\frac{s^2}{2} + o(1). \qquad\qquad (\text{A.171})
\end{aligned}$$

In addition, as $nh/(n-p) \to 0$ and $nh/(n+r-p) \to 0$, we have

$$\begin{aligned}
&\frac{m(n+r-p)}{2}\log\frac{n+r-p}{n+r-p+nh} \qquad\qquad (\text{A.172})\\
= \;& -\frac{m(n+r-p)}{2}\left\{\frac{nh}{n+r-p} - \frac{n^2h^2}{2(n+r-p)^2} + R_{n,1}\right\},
\end{aligned}$$

and

$$\frac{m(n-p)}{2}\log\frac{n-p+nh}{n-p} = \frac{m(n-p)}{2}\left\{\frac{nh}{n-p} - \frac{n^2h^2}{2(n-p)^2} + R_{n,2}\right\}, \qquad \text{(A.173)}$$

where the remainder terms

$$R_{n,1} = \sum_{k=3}^{\infty}\frac{1}{k}(-1)^{k+1}\frac{(nh)^k}{(n+r-p)^k}, \quad R_{n,2} = \sum_{k=3}^{\infty}\frac{1}{k}(-1)^{k+1}\frac{(nh)^k}{(n-p)^k}. \qquad \text{(A.174)}$$

Then we have

$$\begin{aligned} & \text{(A.172)} + \text{(A.173)} \\ = \; & \frac{mn^2h^2}{4(n+r-p)} - \frac{mn^2h^2}{4(n-p)} - \frac{m(n+r-p)}{2}R_{n,1} + \frac{m(n-p)}{2}R_{n,2} \\ = \; & -\frac{s^2}{2} + o(1), \qquad \text{(A.175)} \end{aligned}$$

where in the last equation, we use (A.171) and Lemma A.4.7.

Furthermore, by $nh/(n+r-p) \to 0$ and (A.169), we have

$$\begin{aligned} & \frac{nh}{2}m\log\frac{n-p+nh}{n+r-p+nh} \\ = \; & \frac{nh}{2}m\log\frac{n-p}{n+r-p} + \frac{nmh}{2}\frac{n+r-p}{n-p}\frac{nhr}{(n+r-p)(n+r-p+nh)} + o(1) \\ = \; & \frac{nh}{2}m\log\frac{n-p}{n+r-p} + s^2 + o(1). \qquad \text{(A.176)} \end{aligned}$$

Combining (A.170), (A.171), (A.175) and (A.176), we obtain $\log E\exp\left\{\frac{\log L_n - \mu_n/2}{n\sigma_n/2}s\right\} = s^2/2 + o(1)$.

Case 1.3 When $m/(n+r-p) \to 0$ and $m/(n-p) \to \gamma \in (0,1]$, we know (A.164) still holds following similar analysis to Case 1.1. And (A.167) also holds following similar analysis to Case 1.2. To establish (A.160), we next show that under this case, the difference between the result of (A.165) and (A.167) is ignorable.

$$\begin{aligned}
&(\text{A.167}) - (\text{A.165}) \\
= &\, mnh + \frac{mnh}{2}\left\{\log\left(\frac{n+r-p}{2}\right) - 1\right\} + \frac{n^2h^2}{4} \times \frac{m}{n+r-p} \\
&- \frac{m(n+r-p)}{2}\log\left(1+\frac{nh}{n+r-p}\right) - \frac{mnh}{2}\log\left(\frac{n+r-p}{2}+\frac{nh}{2}\right) + o(1).
\end{aligned} \tag{A.177}$$

We then analyze the terms in (A.177) separately.

Since $m/(n-p) \to \gamma \in (0,1]$, similarly to (A.163), we know that $nh = 2s/\sigma_n = O(s)$. As $m/(n+r-p) \to 0$, it follows that $n^2h^2m/(n+r-p) \to 0$. Applying Taylor expansion, we then have

$$\frac{mnh}{2}\log\left(\frac{n+r-p}{2}+\frac{nh}{2}\right) = \frac{mnh}{2}\log\left(\frac{n+r-p}{2}\right) + o(1). \tag{A.178}$$

Similarly, by $nh = O(s)$, $m/(n+r-p) \to 0$, and Taylor expansion, we have

$$\frac{m(n+r-p)}{2}\log\left(1+\frac{nh}{n+r-p}\right) = \frac{mnh}{2} + o(1). \tag{A.179}$$

In summary, combining (A.178) and (A.179), we have (A.177) = (A.167) − (A.165) = $o(1)$. Then by the results in Case 1.1, we get the same conclusion as in Case 1.1.

Case 2. When $m > r$, $m \to \infty$. According to Lemma A.4.2, we can make the following substitution $m \to r, \quad r \to m, \quad n-p \to n+r-p-m$. And the theorem can be proved following similar analysis when $m \to \infty$, $n-p+r-m \to \infty$.

A.4.4 Lemmas in the proof of Theorem 2.4.3

Lemma A.4.1 (Corollary 10.5.2 in Muirhead (2009)). *Under the null hypothesis, L_n's h-th moment can be written as*

$$\mathrm{E}(L_n^h) = \frac{\Gamma_m\{\frac{1}{2}n(1+h) - \frac{1}{2}p\}\Gamma_m\{\frac{1}{2}(n+r-p)\}}{\Gamma_m\{\frac{1}{2}(n-p)\}\Gamma_m\{\frac{1}{2}n(1+h)+\frac{1}{2}(r-p)\}}.$$

Lemma A.4.2 (Theorem 10.5.3 in Muirhead (2009)). *Under the null hypothesis, when $n-p \geq m$ and $r \geq m$, $\frac{2}{n}\log L_n$ has the same distribution as $\sum_{i=1}^m \log V_i$, where V_i's are independent random variables and $V_i \sim$ beta$(\frac{1}{2}(n-p-i+1), \frac{1}{2}r)$; when $n-p \geq m \geq r$, $\frac{2}{n}\log L_n$ has the same distribution as $\sum_{i=1}^r \log V_i$, where V_i's are independent and $V_i \sim$ beta$(\frac{1}{2}(n+r-p-m-i+1), \frac{1}{2}m)$.*

Lemma A.4.3 (Theorem 2.1.12 in Muirhead (2009)). *The multivariate Gamma function defined in* (A.162) *can be written as* $\Gamma_m(a) = \pi^{m(m-1)/4}\prod_{j=1}^m \Gamma(a-(j-1)/2)$.

Lemma A.4.4. *Consider m is fixed and $a \to \infty$. We have*

$$\frac{1}{a-1}\sum_{i=1}^m \frac{i-1}{a-i} = \Big\{\frac{1}{2}\Big(\sigma_a^2 - \frac{m}{(a-1)^2}\Big)\Big\}\{1+O(1/a)\}, \qquad \text{(A.180)}$$

$$\sum_{i=1}^m \{\log(a-1) - \log(a-i)\} = -\mu_a + O\Big(\frac{m^2}{a^2}\Big), \qquad \text{(A.181)}$$

where $\mu_a = -(m-a+3/2)\log\{1-m/(a-1)\} + (a-1)m/a$ and $\sigma_a^2 = -2[m/(a-1) + \log\{1-m/(a-1)\}]$.

Proof. We first prove (A.180). As m is fixed and $a \to \infty$, we have

$$\sigma_a^2 = -2\Big[\frac{m}{a-1} + \log\Big(1-\frac{m}{a-1}\Big)\Big] = \Big(\frac{m}{a-1}\Big)^2\{1+O(m/a)\},$$

and

$$\frac{1}{a-1}\sum_{i=1}^{m}\frac{i-1}{a-i}=\frac{1}{a-1}\sum_{i=1}^{m}\frac{i-1}{a-1}+\epsilon_a=\frac{m(m-1)}{2(a-1)^2}+\epsilon_a,$$

where $|\epsilon_a|\leq 2(a-1)^{-3}\sum_{i=1}^{m}(i-1)^2\leq 3(m/a)^3$. Therefore,

$$\frac{1}{a-1}\sum_{i=1}^{m}\frac{i-1}{a-i}=\frac{m(m-1)}{2(a-1)^2}\Big\{1+O\Big(\frac{m}{a}\Big)\Big\}=\Big[\frac{1}{2}\Big\{\sigma_a^2-\frac{m}{(a-1)^2}\Big\}\Big]\{1+O(1/a)\},$$

where in the last equation, we use the fact that $O(m/a)=O(1/a)$ as m is fixed. Then (A.180) is proved.

We then prove (A.181). Recall Stirling's formula, (see, e.g., p. 368 Gamelin, 2001) $\log\Gamma(x)=(x-1/2)\log x-x+\log\sqrt{2\pi}+1/(12x)+O(x^{-3})$ as $x\to\infty$. Therefore,

$$\begin{aligned}
&\log\Gamma(a-1)-\log\Gamma(a-m-1)\\
=\;&(a-3/2)\log(a-1)-(a-m-3/2)\log(a-m-1)-m\\
&+\frac{1}{12}\Big(\frac{1}{a-1}-\frac{1}{a-m-1}\Big)+O(a^{-3})\\
=\;&(a-3/2)\log(a-1)-(a-m-3/2)\log(a-m-1)-m+O(ma^{-2}).
\end{aligned}$$

Since for integers $k\geq 1$, $\Gamma(k)=(k-1)!=\Pi_{i=1}^{k-1}i$. Then we have

$$\begin{aligned}
&\sum_{i=1}^{m}\{\log(a-1)-\log(a-i)\}\\
=\;&m\log(a-1)-\{\log\Gamma(a-1)-\log\Gamma(a-m-1)\}+\log\Big(1-\frac{m}{a-1}\Big)\\
=\;&m\log(a-1)-(a-3/2)\log(a-1)+(a-m-3/2)\log(a-m-1)\\
&+m+\log\Big(1-\frac{m}{a-1}\Big)+O(ma^{-2})\\
=\;&-(m-a+3/2)\log\Big(1-\frac{m}{a-1}\Big)+\frac{a-1}{a}m+O\Big(\frac{m^2}{a^2}\Big)\\
=\;&-\mu_a+O\Big(\frac{m^2}{a^2}\Big),
\end{aligned}$$

where we use the fact that $\frac{a-2}{a-1}m = \frac{a-1}{a}m + O(ma^{-2})$. □

Lemma A.4.5. *Consider m is fixed and $a \to \infty$. Define*

$$g_i(x) = \Big(\frac{a-i}{2} + x\Big) \log \Big(\frac{a-i}{2} + x\Big) - \Big(\frac{a-1}{2} + x\Big) \log \Big(\frac{a-1}{2} + x\Big)$$

for $1 \le i \le m$ and $x > -(a-m)/2$. Let μ_a and σ_a be as in Lemma A.4.4. If $t = o(a)$ and $mt^2/a^2 = o(1)$, we have that as $a \to \infty$, $\sum_{i=1}^m \{g_i(t) - g_i(0)\} = \mu_a t + \sigma_a^2 t^2/2 + o(1)$.

Proof. We know for $1 \le i \le m$,

$$\begin{aligned}
g_i'(x) &= \log(\frac{a-i}{2} + x) - \log(\frac{a-1}{2} + x), \\
g_i''(x) &= \frac{1}{\frac{a-i}{2} + x} - \frac{1}{\frac{a-1}{2} + x} = \frac{\frac{i-1}{2}}{(\frac{a-i}{2} + x)(\frac{a-1}{2} + x)}, \\
g_i^{(3)}(x) &= -\frac{1}{(\frac{a-i}{2} + x)^2} + \frac{1}{(\frac{a-1}{2} + x)^2} \\
&= -\frac{\frac{i-1}{2} \cdot \frac{2a-i-1}{2} + (i-1)x}{(\frac{a-i}{2} + x)^2(\frac{a-1}{2} + x)^2}.
\end{aligned}$$

By Taylor expansion,

$$\begin{aligned}
g_i(t) - g_i(0) &= g_i'(0)t + \frac{t^2}{2} g_i''(0) + \frac{t^3}{6} g_i^{(3)}(\xi_i) \\
&= \{\log(a-i) - \log(a-1)\}t + \frac{i-1}{(a-1)(a-i)} t^2 + \frac{t^3}{6} g_i^{(3)}(\xi_i).
\end{aligned}$$

For $1 \le i \le m$, fixed m and $0 \le \xi_i \le t = o(a)$, we have $\sup_{|\xi_i| \le |t|, 1 \le i \le m} |g_i^{(3)}(\xi_i)| \le ca^{-3}$, where c denotes an universal constant. Therefore, as $t = o(a)$, $|t^3 g_i^{(3)}(\xi_i)| \le ct^3 a^{-3} = o(1)$. In addition, by Lemma A.4.4, and the fact that $mt^2/(a-1)^2 = o(1)$, we have as $a \to \infty$,

$$\sum_{i=1}^m \{g_i(t) - g_i(0)\} = \mu_a t + \Big[\frac{1}{2}\Big(\sigma_a^2 - \frac{m}{(a-1)^2}\Big)\Big] t^2 + o(1) = \mu_a t + \frac{\sigma_a^2}{2} t^2 + o(1).$$

□

Lemma A.4.6. *Consider $n-p \to \infty$, $r \to \infty$, $m/(n-p) \to 0$ and $m/(n-p+r) \to 0$. For $t = nh/2$, $a = n-p+r$ or $a = n-p$, we have*

$$\log \frac{\Gamma_m(\frac{a-1}{2}+t)}{\Gamma_m(\frac{a-1}{2})} = \upsilon_a t + \vartheta_a t^2 + \gamma_a(t) + o(1),$$

where

$$\begin{aligned}
\upsilon_a &= -\left[2m + (a-m-3/2)\log\{1-m/(a-1)\}\right];\\
\vartheta_a &= -\left[m/(a-1) + \log\{1-m/(a-1)\}\right];\\
\gamma_a(t) &= m\Big\{\Big(\frac{a-1}{2}+t\Big)\log\Big(\frac{a-1}{2}+t\Big) - \frac{a-1}{2}\log\Big(\frac{a-1}{2}\Big)\Big\}.
\end{aligned}$$

Proof. By Lemma A.4.3, we know

$$\log \frac{\Gamma_m(\frac{a-1}{2}+t)}{\Gamma_m(\frac{a-1}{2})} = \sum_{i=1}^{m} \log \frac{\Gamma(\frac{a-i}{2}+t)}{\Gamma(\frac{a-i}{2})}. \tag{A.182}$$

To prove the lemma, we expand each summed term in (A.182), $\log\{\Gamma(\frac{a-i}{2}+t)/\Gamma(\frac{a-i}{2})\}$, by Lemma A.1. in Jiang and Qi (2015). To apply the lemma, we first need to check the condition that for each $1 \le i \le m$, $t \in [-\delta(a-i)/2, \delta(a-i)/2]$ for any given $\delta \in (0,1)$.

Recall that we previously define $nh = 2s/\sigma_n$ in Section A.4.3. Then $t = nh/2 = s\sigma_n^{-1}$. Note that when $m/(n-p)$ and $m/(n-p+r) \to 0$,

$$\sigma_n^2 = \frac{1}{2}\log\left(1-\frac{m}{n+r-p}\right) - \frac{1}{2}\log\left(1-\frac{m}{n-p}\right) = \frac{mr}{2(n-p)(n+r-p)}\{1+o(1)\}.$$

Thus we have

$$t = O(s)\sqrt{\frac{(n-p)(n-p+r)}{mr}}. \tag{A.183}$$

For $a = n - p + r$ or $a = n - p$, and $1 \leq i \leq m$, by (A.183), we then have

$$\begin{aligned} \frac{t}{a-i} &\leq \frac{t}{n-p-m} = O(s)\sqrt{\frac{(n-p)(n-p+r)}{mr(n-p-m)^2}} \\ &= O(s)\sqrt{\left\{\frac{1}{mr} + \frac{1}{m(n-p)}\right\}\{1+o(1)\}} = o(1), \end{aligned}$$

where the last two equations follow from the condition that $m/(n-p) \to 0, r \to \infty$ and $n - p \to \infty$. Then we know that for each $1 \leq i \leq m$, $t \in [-\delta(a-i)/2, \delta(a-i)/2]$ for any given $\delta \in (0, 1)$.

Therefore, the condition of Lemma A.1. in Jiang and Qi (2015) is satisfied. By that lemma, we know when $a \to \infty$, for uniformly $1 \leq i \leq m$,

$$\log \frac{\Gamma(\frac{a-i}{2} + t)}{\Gamma(\frac{a-i}{2})} = \left(\frac{a-i}{2} + t\right) \log\left(\frac{a-i}{2} + t\right) - \frac{a-i}{2}\log\frac{a-i}{2} - t - \frac{t}{a-i} + O\left(\frac{t^2}{a^2}\right).$$

Write $\frac{t}{a-i} = \frac{t}{a} + \frac{t}{a} \times \frac{i}{a-i}$. Then similarly to Lemma A.4.4, we have

$$\sum_{i=1}^{m} \frac{t}{a-i} = \frac{mt}{a} + \frac{tm(m+1)}{2a(a-1)} + O\left\{\frac{t}{a} \times \left(\frac{m}{a}\right)^3\right\}. \tag{A.184}$$

For $a = n - p$, by (A.183), $m/(n-p) \to 0$ and $m \leq r$,

$$\begin{aligned} \frac{tm(m+1)}{a(a-1)} &= O(s)\sqrt{\frac{(n-p)(n-p+r)}{mr}}\frac{m^2}{(n-p)^2} \\ &= O(s)\sqrt{\frac{m}{\min\{n-p, r\}}}\frac{m}{n-p} = o(1). \end{aligned}$$

For $a = n - p + r$, similar conclusion, $tm(m+1)/\{a(a-1)\} = o(1)$, holds by substituting $n - p$ with $n - p + r$. In addition, for $a = n - p$ or $a = n - p + r$, by (A.183),

$$\frac{t}{a} \leq \frac{t}{n-p} = O(s)\sqrt{\frac{\max\{n-p, r\}}{mr(n-p)}} = O(s)\sqrt{\frac{1}{m \times \min\{n-p, r\}}} = o(1). \tag{A.185}$$

Then based on (A.184) and (A.185), we obtain

$$\sum_{i=1}^{m}\Big\{-t-\frac{t}{a-i}+O(t^2/a^2)\Big\}=-mt-\frac{mt}{a}+o(1).$$

Therefore, from (A.182), we have

$$\begin{aligned}&\log\frac{\Gamma_m(\frac{a-1}{2}+t)}{\Gamma_m(\frac{a-1}{2})} &(\text{A.186})\\ &=-\frac{(a+1)mt}{a}+\sum_{i=1}^{m}\Big\{\Big(\frac{a-i}{2}+t\Big)\log\Big(\frac{a-i}{2}+t\Big)-\frac{a-i}{2}\log\frac{a-i}{2}\Big\}+o(1).\end{aligned}$$

For $1\le i\le m$, define the function

$$g_i(x)=\Big(\frac{a-i}{2}+x\Big)\log\Big(\frac{a-i}{2}+x\Big)-\Big(\frac{a-1}{2}+x\Big)\log\Big(\frac{a-1}{2}+x\Big),$$

and $x>-(a-m)/2$. We then know that the summation term "$\sum$" in (A.186) equals to

$$m\Big[\Big(\frac{a-1}{2}+t\Big)\log\Big(\frac{a-1}{2}+t\Big)-\frac{a-1}{2}\log\frac{a-1}{2}\Big]+\sum_{i=1}^{m}\{g_i(t)-g_i(0)\}. \quad (\text{A.187})$$

We then examine the function $\sum_{i=1}^{m}\{g_i(t)-g_i(0)\}$ in (A.187). Note that by (A.185), we know $t=o(a)$, $mt^2/a^2=o(1)$ and $mt/a=O(1)$ as $m<n-p$ and $m\le r$. Thus the conditions of Lemma A.4.5 and Lemma A.3. in Jiang and Qi (2015) are satisfied when m is fixed and $m\to\infty$ respectively. When m is fixed, we apply Lemma A.4.5; when $m\to\infty$, we apply Lemma A.3. in Jiang and Qi (2015). Then we obtain $\sum_{i=1}^{m}\{g_i(t)-g_i(0)\}=\mu_a t+\sigma_a^2t^2/2+o(1)$, where

$$\mu_a=(m-a+3/2)\log\Big(1-\frac{m}{a-1}\Big)-m\frac{a-1}{a},\quad \sigma_a^2=-2\Big[\frac{m}{a-1}+\log\Big(1-\frac{m}{a-1}\Big)\Big].$$

Therefore, the proposition can be proved by noticing

$$\begin{aligned}
\upsilon_a &= -\frac{(a+1)m}{a} + \mu_a; \quad \vartheta_a = \sigma_a^2/2;\\
\gamma_a(t) &= m\Big[\Big(\frac{a-1}{2}+t\Big)\log\Big(\frac{a-1}{2}+t\Big) - \frac{a-1}{2}\log\frac{a-1}{2}\Big].
\end{aligned}$$

□

Lemma A.4.7. *Under Case 1 in Section A.4.3, $R_{n,1}$ and $R_{n,2}$ defined in* (A.174) *satisfy $-m(n+r-p)R_{n,1}/2 + m(n-p)R_{n,2}/2 = o(1)$.*

Proof. Note that

$$\begin{aligned}
& -\frac{m(n+r-p)}{2}R_{n,1} + \frac{m(n-p)}{2}R_{n,2}\\
=\; & \frac{m}{2}\Big[\sum_{k=3}^{\infty}\frac{1}{k}(-nh)^k\Big\{\frac{1}{(n+r-p)^{k-1}} - \frac{1}{(n-p)^{k-1}}\Big\}\Big]\\
=\; & \frac{mnh}{2}\sum_{k=3}^{\infty}\frac{1}{k}\Big(\frac{-nh}{n+r-p}\Big)^{k-1}\sum_{q=1}^{k-1}\binom{k-1}{q}\Big(\frac{r}{n-p}\Big)^q. \qquad \text{(A.188)}
\end{aligned}$$

If $r/(n-p) = 1$,

$$|(\text{A.188})| \le mnh\sum_{k=3}^{\infty}\Big(\frac{2nh}{n+r-p}\Big)^{k-1} = O\Big\{mnh\frac{n^2h^2}{(n+r-p)^2}\Big\},$$

where

$$\begin{aligned}
\frac{mnh \times n^2h^2}{(n+r-p)^2} &= O\Big\{\frac{m\sqrt{(n-p)(n+r-p)}}{\sqrt{mr}} \times \frac{(n-p)(n+r-p)}{mr(n+r-p)^2}\Big\}\\
&= O\Big\{\frac{m\sqrt{r^2}}{\sqrt{mr}} \times \frac{r^2}{mr \times r^2}\Big\} = o(1),
\end{aligned}$$

as $r \to \infty$.

If $r/(n-p) > 1$, as $\{nh/(n+r-p)\}\times\{r/(n-p)\} = O\{\sqrt{r}/\sqrt{m(n-p)(n+r-p)}\} =$

$o(1)$,

$$|(\text{A.188})| \leq mnh\sum_{k=3}^{\infty}\Big(\frac{2nh}{n+r-p}\times\frac{r}{n-p}\Big)^{k-1} = O\Big\{mnh\Big(\frac{2nh}{n+r-p}\Big)^2\Big(\frac{r}{n-p}\Big)^2\Big\},$$

where

$$\begin{aligned} & mnh\Big(\frac{nh}{n+r-p}\Big)^2\Big(\frac{r}{n-p}\Big)^2 \\ = \;& O\Big\{\frac{m\sqrt{(n-p)(n+r-p)}}{\sqrt{mr}}\times\frac{(n-p)(n+r-p)}{mr(n+r-p)^2}\times\frac{r^2}{(n-p)^2}\Big\} \\ = \;& O\Big\{\frac{r}{\sqrt{mr(n+r-p)(n-p)}}\Big\} = o(1), \end{aligned}$$

as $n+r-p \geq r$ and $n-p \to \infty$.

If $r/(n-p) < 1$,

$$|(\text{A.188})| \leq mnh\sum_{k=3}^{\infty}\Big(\frac{nh}{n+r-p}\Big)^{k-1}\frac{r}{(n-p)} = O\Big\{mnh\frac{(nh)^2}{(n+r-p)^2}\times\frac{r}{(n-p)}\Big\},$$

where

$$\begin{aligned} & mnh\frac{(nh)^2}{(n+r-p)^2}\times\frac{r}{(n-p)} \\ = \;& O\Big\{\frac{m\sqrt{(n-p)(n+r-p)}}{\sqrt{mr}}\times\frac{(n-p)(n+r-p)}{mr(n+r-p)^2}\times\frac{r}{(n-p)}\Big\} \\ = \;& O\Big\{\frac{\sqrt{n-p}}{\sqrt{mr(n+r-p)}}\Big\} = o(1). \end{aligned}$$

□

APPENDIX B

Appendix of Chapter III

This appendix is for Chapter III and is organized as follows. Sections B.1–B.4 present proofs of theoretical results in Sections 3.2–3.5. Section B.5 proves all the technical lemmas used in the Appendix B. Section B.6 discusses the computation of the U-Statistics in Chapter III. Section B.7 provides supplementary simulations for Chapter III.

B.1 Proofs of Theoretical Results in Section 3.2

B.1.1 Proof of Theorem 3.2.1

For the covariance testing example in Section 3.2, $\mathcal{U}(a)$ is location invariant by Property 3.2.1, and $\mathcal{U}(\infty)$ is also location invariant straightforwardly by its expression in (3.9). Then we assume without loss of generality that $\mathrm{E}(\mathbf{x}) = \mathbf{0}$ in this section. To prove Theorem 3.2.1, we first derive the variances and the covariances of the U-statistics, and then prove the asymptotic joint normality of the U-statistics.

In particular, the following Lemma B.1.1 derives the asymptotic form of variance $\sigma^2(a)$ in (3.8).

Lemma B.1.1. *Under the conditions of Theorem 3.2.1, for any finite integer a,* *following the notation in* (3.6),

$$\sigma^2(a) = \frac{a!}{P_a^n} \sum_{\substack{1 \le j_1 \ne j_2 \le p; \\ 1 \le j_3 \ne j_4 \le p}} (\Pi_{j_1,j_2,j_3,j_4})^a \{1 + o(1)\},$$

which is of order $\Theta(p^2 n^{-a})$. *In addition, for* $\tilde{\mathcal{U}}(a)$ *defined in* (3.5) *and* $\tilde{\mathcal{U}}^*(a) := \mathcal{U}(a) - \tilde{\mathcal{U}}(a)$, *we have* $\mathrm{var}\{\mathcal{U}(a)\} = \mathrm{var}\{\tilde{\mathcal{U}}(a)\}\{1 + o(1)\}$, $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1) \times \mathrm{var}\{\tilde{\mathcal{U}}(a)\}$, *and* $\tilde{\mathcal{U}}^*(a)/\sigma(a) \xrightarrow{P} 0$.

Proof. *See Section B.5.2 on Page 286.* □

Moreover, the following Lemma B.1.2 shows that the covariances between different $\mathcal{U}(a)$'s asymptotically converge to 0.

Lemma B.1.2. *Under the conditions of Theorem 3.2.1, for finite integers* $a \ne b$, $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} \to 0$, *as* $n, p \to \infty$.

Proof. *See Section B.5.3 on Page 305.* □

Lemmas B.1.1 and B.1.2 together show that the covariance matrix of the U-statistics $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ converges to I_m asymptotically. To finish the proof of Theorem 3.2.1, it remains to show that the joint limiting distribution of the U-statistics is normal.

For finite integers $a_1, \ldots, a_m$, to obtain the joint asymptotic normality of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$, by the Cramér-Wold theorem, it is equivalent to prove that any fixed linear combination of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ converges to normal. Recall that Lemma B.1.1 shows that $\tilde{\mathcal{U}}^*(a)/\sigma(a) \xrightarrow{P} 0$ for any finite integer a. Thus by Slutsky's theorem, it suffices to prove that any fixed linear combination of $[\tilde{\mathcal{U}}(a_1)/\sigma(a_1), \ldots, \tilde{\mathcal{U}}(a_m)/\sigma(a_m)]^\intercal$ converges to normal. To be specific, we show that for constants $t_1, \ldots, t_m$ satisfying $\sum_{r=1}^m t_r^2 = 1$,

$$Z_n := \sum_{r=1}^m t_r \tilde{\mathcal{U}}(a_r)/\sigma(a_r) \xrightarrow{D} \mathcal{N}(0, 1). \tag{B.1}$$

To prove (B.1), we apply the martingale central limit theorem in Heyde and Brown (1970) (similar arguments can date back to Bai and Saranadasa (1996)). Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_k = \sigma\{\mathbf{x}_1, \cdots, \mathbf{x}_k\}$, and $\mathrm{E}_k(\cdot)$ denote the conditional expectation given $\mathcal{F}_k$ for $k = 1, \cdots, n$. Define $D_{n,k} = (\mathrm{E}_k - \mathrm{E}_{k-1})Z_n$ and $\pi_{n,k}^2 = \mathrm{E}_{k-1}(D_{n,k}^2)$. Note that $\mathrm{E}_0(\cdot) = \mathrm{E}(\cdot)$, and $\mathrm{E}(Z_n) = 0$ as $\mathrm{E}(\mathbf{x}) = \mathbf{0}$. It follows that $Z_n = \sum_{k=1}^n D_{n,k}$. By martingale central limit theorem, to prove (B.1), it is sufficient to show

$$\sum_{k=1}^{n} \pi_{n,k}^2/\mathrm{var}(Z_n) \xrightarrow{P} 1, \qquad \sum_{k=1}^{n} \mathrm{E}(D_{n,k}^4)/\mathrm{var}^2(Z_n) \to 0. \tag{B.2}$$

Here $\mathrm{var}(Z_n) \to \sum_{r=1}^m t_r^2 = 1$ by Lemmas B.1.1 and B.1.2, and $\mathrm{E}(\sum_{k=1}^n \pi_{n,k}^2) = \mathrm{var}(Z_n)$ by the following Lemma B.1.3.

Lemma B.1.3. *Under the conditions of Theorem 3.2.1,* $\mathrm{E}(\sum_{k=1}^n \pi_{n,k}^2) = \mathrm{var}(Z_n)$.

Proof. *See Section B.5.4 on Page 305.* □

Therefore to prove (B.2), it suffices to show

$$\mathrm{var}\left(\sum_{k=1}^{n} \pi_{n,k}^2\right) \to 0 \quad \text{and} \quad \sum_{k=1}^{n} \mathrm{E}(D_{n,k}^4) \to 0. \tag{B.3}$$

Note that $D_{n,k}$ and $\pi_{n,k}^2$ in (B.3) can be written as $D_{n,k} = \sum_{r=1}^m t_r A_{n,k,a_r}$ and $\pi_{n,k}^2 = \sum_{1 \le r_1, r_2 \le m} \mathrm{E}_{k-1}(A_{n,k,a_{r_1}} A_{n,k,a_{r_2}})$, where we define $A_{n,k,a} = (\mathrm{E}_k - \mathrm{E}_{k-1})\{\tilde{\mathcal{U}}(a)/\sigma(a)\}$ for each finite integer a. The following Lemma B.1.4 gives the explicit form of $A_{n,k,a}$.

Lemma B.1.4. *For finite integer a, when $k < a$, $A_{n,k,a} = 0$; when $k \ge a$,*

$$A_{n,k,a} = \frac{a}{\sigma(a) P_a^n} \sum_{1 \le i_1 \ne \cdots \ne i_{a-1} \le k-1} \ \sum_{1 \le j_1 \ne j_2 \le p} (x_{k,j_1} x_{k,j_2}) \times \prod_{t=1}^{a-1} (x_{i_t,j_1} x_{i_t,j_2}).$$

Proof. *See Section B.5.5 on Page 306.* □

With the form of $A_{n,k,a}$ in Lemma B.1.4, the forms of $D_{n,k}$ and $\pi_{n,k}^2$ can be obtained, and we can prove the next two Lemmas B.1.5 and B.1.6, which suggest that (B.3) holds.

Lemma B.1.5. *Under the conditions of Theorem 3.2.1,* $\mathrm{var}(\sum_{k=1}^{n}\pi_{n,k}^2) \to 0$*. In particular, under Condition 3.2.2,* $\mathrm{var}(\sum_{k=1}^{n}\pi_{n,k}^2) = O(p^{-1}\log^3 p)$*; under Condition 3.2.2*,* $\mathrm{var}(\sum_{k=1}^{n}\pi_{n,k}^2) = O(n^{-1}+p^{-2})$*.*

Proof. *See Section B.5.6 on Page 308.* □

Lemma B.1.6. *Under the conditions of Theorem 3.2.1,* $\sum_{k=1}^{n}\mathrm{E}(D_{n,k}^4) = O(1/n)$*.*

Proof. *See Section B.5.7 on Page 327.* □

Finally, by Heyde and Brown (1970), we have as $n, p \to \infty$,

$$\begin{aligned}
&\sup_t \Big| P(Z_n \le t) - \Phi(t) \Big| \qquad\qquad (B.4)\\
\le\ & C\left\{\mathrm{E}\left[\frac{\sum_{k=1}^{n}\mathrm{E}_{k-1}(D_{n,k}^2)}{\mathrm{var}(Z_n)} - 1\right]^2 + \frac{\sum_{k=1}^{n}\mathrm{E}\left(D_{n,k}^4\right)}{\mathrm{var}^2(Z_n)}\right\}^{1/5}\\
\to\ & 0,
\end{aligned}$$

which proves (B.1). In summary, Theorem 3.2.1 is proved.

B.1.2 Proof of Theorem 3.2.3

In this section, we first introduce some notation, and then present the proof.

Notation. For $\mathcal{U}(a)$ in (3.3), by the symmetricity of covariance matrix, we can replace $\sum_{1\le j_1\ne j_2\le p}$ by $2\times\sum_{1\le j_1<j_2\le p}$. This implies that the summation over $\{(j_1,j_2): 1\le j_1\ne j_2\le p\}$ is equivalent to the summation over $\{(j_1,j_2): 1\le j_1<j_2\le p\}$ up to a constant. Without loss of generality, we consider $j_1<j_2$ below. We rewrite the index set $\{(j_1,j_2): 1\le j_1<j_2\le p\}$ as

$$L := \Big\{(j_l^1, j_l^2) : 1 \le l \le q = \binom{p}{2}\Big\}, \qquad (B.5)$$

where $j_l^1 = \arg\min_{1\le k\le p-1}\{\sum_{t=1}^{k}(p-t)\ge l\}$ and $j_l^2 = l + j_l^1 - \sum_{t=1}^{j_l^1-1}(p-t)$. For each $(j_l^1, j_l^2)\in L$, define

$$U_l^a = \sum_{1\le i_1\ne\ldots\ne i_a\le n}\prod_{k=1}^{a} x_{i_k,j_l^1}x_{i_k,j_l^2}. \tag{B.6}$$

Then $\tilde{\mathcal{U}}(a) = 2(P_a^n)^{-1}\sum_{l=1}^{q} U_l^a$ following the definition in (3.5). Furthermore, we define

$$\begin{aligned}
\tilde{G}_l &= \sum_{i=1}^{n}\frac{x_{i,j_l^1}}{\sqrt{\sigma_{j_l^1,j_l^1}}}\times\frac{x_{i,j_l^2}}{\sqrt{\sigma_{j_l^2,j_l^2}}}, \qquad\qquad (\text{B.7})\\
M_n &= \max_{1\le l\le q}(\tilde{G}_l)^2,\\
\hat{G}_l &= \sum_{i=1}^{n}\frac{x_{i,j_l^1}}{\sqrt{\sigma_{j_l^1,j_l^1}}}\times\frac{x_{i,j_l^2}}{\sqrt{\sigma_{j_l^2,j_l^2}}}\mathbf{1}\Big\{\Big|\frac{x_{i,j_l^1}}{\sqrt{\sigma_{j_l^1,j_l^1}}}\times\frac{x_{i,j_l^2}}{\sqrt{\sigma_{j_l^2,j_l^2}}}\Big|\le\tau_n\Big\}\\
&\quad -\mathrm{E}\Bigg[\sum_{i=1}^{n}\frac{x_{i,j_l^1}}{\sqrt{\sigma_{j_l^1,j_l^1}}}\times\frac{x_{i,j_l^2}}{\sqrt{\sigma_{j_l^2,j_l^2}}}\mathbf{1}\Big\{\Big|\frac{x_{i,j_l^1}}{\sqrt{\sigma_{j_l^1,j_l^1}}}\times\frac{x_{i,j_l^2}}{\sqrt{\sigma_{j_l^2,j_l^2}}}\Big|\le\tau_n\Big\}\Bigg],\\
\hat{M}_n &= \max_{1\le l\le q}(\hat{G}_l)^2,
\end{aligned}$$

where we define $\sigma_{j_l^1,j_l^1} = \mathrm{var}(x_{i,j_l^1})$, $\sigma_{j_l^2,j_l^2} = \mathrm{var}(x_{i,j_l^2})$, $\tau_n = \tau\log(p+n)$ with τ being a sufficiently large positive constant and $\mathbf{1}\{\cdot\}$ represents an indicator function. In addition, we define $|\mathbf{a}|_{\min} = \min_{1\le i\le p}|a_i|$ for $\mathbf{a}\in\mathbb{R}^p$, and

$$y_p = 4\log p - \log\log p + y. \tag{B.8}$$

Proof. Similarly to Section B.1.1, since $\mathcal{U}(a)$ in (3.3) and $\mathcal{U}(\infty)$ in (3.9) are location invariant, we assume without loss of generality that $\mathrm{E}(\mathbf{x}) = \mathbf{0}$.

To prove Theorem 3.2.3, we first establish the asymptotic independence between $\hat{M}_n/n$ and $\tilde{\mathcal{U}}(a)/\sigma(a_r)$ for $r = 1,\ldots,m$, and then we show that $n\mathcal{U}^2(\infty)$ and $\mathcal{U}(a_r)$ are close to $\hat{M}_n/n$ and $\tilde{\mathcal{U}}(a_r)$, respectively. Specifically, the following Lemma B.1.7 shows that $\hat{M}_n/n$ and $\tilde{\mathcal{U}}(a_r)/\sigma(a_r)$'s are asymptotically independent.

Lemma B.1.7. *Under the conditions of Theorem 3.2.3, when* $\tau > 0$ *in* (B.7) *is a sufficiently large constant,*

$$\Big| P\Big(\frac{\hat{M}_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a_1)}{\sigma(a_1)} \le z_1, \dots, \frac{\tilde{\mathcal{U}}(a_m)}{\sigma(a_m)} \le z_m\Big) \\ - P\Big(\frac{\hat{M}_n}{n} > y_p\Big) \prod_{r=1}^{m} P\Big(\frac{\tilde{\mathcal{U}}(a_r)}{\sigma(a_r)} \le z_r\Big) \Big| \to 0.$$

Proof. *See Section B.5.8 on Page 336.* □

To show that $\hat{M}_n/n$ and $n\mathcal{U}(\infty)^2$ are close, we use M_n/n defined in (B.7) as an intermediate variable. We next prove that M_n/n and $\hat{M}_n/n$ have small difference in the sense that the conclusion in Lemma B.1.7 still holds by replacing $\hat{M}_n$ with M_n. This is formally stated in the following Lemma B.1.8.

Lemma B.1.8. *Under the conditions of Theorem 3.2.3,*

$$\Big| P\Big(\frac{M_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a_1)}{\sigma(a_1)} \le z_1, \dots, \frac{\tilde{\mathcal{U}}(a_m)}{\sigma(a_m)} \le z_m\Big) \\ - P\Big(\frac{M_n}{n} > y_p\Big) \prod_{r=1}^{m} P\Big(\frac{\tilde{\mathcal{U}}(a_r)}{\sigma(a_r)} \le z_r\Big) \Big| \to 0.$$

Proof. *See Section B.5.9 on Page 348.* □

Given Lemma B.1.8, we further prove that M_n/n and $\tilde{\mathcal{U}}(a)/\sigma(a_r)$ are close to $n\mathcal{U}^2(\infty)$ and $\mathcal{U}(a_r)$, respectively. In particular, by the proof of Theorem 3 in Cai and Jiang (2011), we know $\{n^2\mathcal{U}^2(\infty) - M_n\}/n \xrightarrow{P} 0$. In addition, Lemma B.1.1 proves that $\{\mathcal{U}(a_r) - \tilde{\mathcal{U}}(a_r)\}/\sigma(a_r) \xrightarrow{P} 0$. Based on these results and Lemma B.1.8, the following Lemma B.1.9 shows that the conclusion in Lemma B.1.8 still holds by replacing M_n/n with $n\mathcal{U}^2(\infty)$ and replacing $\tilde{\mathcal{U}}(a_r)$ with $\mathcal{U}(a_r)$.

Lemma B.1.9. *Under the conditions of Theorem 3.2.3,*

$$\Big|P\Big(n\mathcal{U}^2(\infty) > y_p, \frac{\mathcal{U}(a_1)}{\sigma(a_1)} \leq z_1, \ldots, \frac{\mathcal{U}(a_m)}{\sigma(a_m)} \leq z_m\Big) - P\Big(n\mathcal{U}^2(\infty) > y_p\Big) \prod_{r=1}^{m} P\Big(\frac{\mathcal{U}(a_r)}{\sigma(a_r)} \leq z_r\Big)\Big| \to 0.$$

Proof. *See Section B.5.10 on Page 350.* □

Lemma B.1.9 then proves Theorem 3.2.3.

B.1.3 Proof of Theorem 3.2.4

As both $\mathcal{U}(a)$ and $\mathbb{V}_u(a)$ are location invariant in the sense of Property 3.2.1, we assume $\mathrm{E}(\mathbf{x}) = \mathbf{0}$. To prove Theorem 3.2.4, we decompose $\mathbb{V}_u(a) = \mathbb{V}_{u,1}(a) + \mathbb{V}_{u,2}(a)$, where we define

$$\mathbb{V}_{u,1}(a) = \frac{2a!}{(P_a^n)^2} \sum_{1\leq j_1 \neq j_2 \leq p} \sum_{1\leq i_1 \neq \ldots \neq i_a \leq n} \prod_{t=1}^{a} x_{i_t,j_1}^2 x_{i_t,j_2}^2,$$

and $\mathbb{V}_{u,2}(a) = \mathbb{V}_u(a) - \mathbb{V}_{u,1}(a)$. The next Lemma B.1.10 shows that $\mathbb{V}_{u,1}(a)$ is of a larger order than $\mathbb{V}_{u,2}(a)$, and thus it is the leading term in $\mathbb{V}_u(a)$.

Lemma B.1.10. *Under the conditions of Theorem 3.2.4,* $\mathbb{V}_{u,1}(a)/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \xrightarrow{P} 1$ *and* $\mathbb{V}_{u,2}(a)/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \xrightarrow{P} 0$.

Proof. *See Section B.5.11 on Page 354.* □

Lemma B.1.10 implies that $\mathbb{V}_u(a)/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \xrightarrow{P} 1$. As $\mathbb{V}_u(a) > 0$ with probability 1, $\mathrm{E}\{\mathbb{V}_{u,1}(a)\}/\mathbb{V}_u(a) \xrightarrow{P} 1$. In addition, note that $\mathrm{E}\{\mathbb{V}_{u,1}(a)\} = 2a!(P_a^n)^{-1} \times \sum_{1\leq j_1\neq j_2\leq p}\{\mathrm{E}(x_{1,j_1}^2 x_{1,j_2}^2)\}^a$. By (B.51) and (B.60) in Section B.5.2, we obtain that $\mathrm{var}\{\mathcal{U}(a)\}/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \to 1$. Therefore,

$$\frac{\mathbb{V}_u(a)}{\mathrm{var}\{\mathcal{U}(a)\}} = \frac{\mathbb{V}_u(a)}{\mathrm{E}\{\mathbb{V}_{u,1}(a)\}} \times \frac{\mathrm{E}\{\mathbb{V}_{u,1}(a)\}}{\mathrm{var}\{\mathcal{U}(a)\}} \xrightarrow{P} 1.$$

B.1.4 Proof of Theorem 3.2.5

Note that Condition 3.2.6 can be viewed as an extension of Condition 3.2.2* to the alternative settings. To be consistent with the notation in Condition 3.2.2*, we let κ_1 denote the constant $\tilde{\kappa}_4$ in Condition 3.2.2* in this proof. We next introduce some notation and then provide the proof.

Notation. For each given $j_1 \in \{1, \ldots, p\}$, we define

$$J_{j_1} = \{(j_1, j_2) : \sigma_{j_1,j_2} \neq 0, 1 \leq j_1 \neq j_2 \leq p\},$$
$$J_{j_1}^c = \{(j_1, j_2) : \sigma_{j_1,j_2} = 0, 1 \leq j_1 \neq j_2 \leq p\}.$$

Then $J_A = \cup_{j_1=1}^p J_{j_1}$, and we correspondingly define $J_A^c = \cup_{j_1=1}^p J_{j_1}^c$, which is the set difference of $\{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p\}$ and J_A. Moreover, we define $F(a, c) = (-1)^c \binom{a}{c} / P^n_{a+c}$, and

$$K(c, j_1, j_2) = F(a, c) \sum_{1 \leq i_1 \neq \ldots \neq i_{a+c} \leq n} \prod_{t=1}^{a-c} (x_{i_t,j_1} x_{i_t,j_2}) \prod_{t=a-c+1}^{a} x_{i_t,j_1} \prod_{t=a+1}^{a+c} x_{i_t,j_2}.$$

We decompose $\mathcal{U}(a) = T_{U,a,1,1} + T_{U,a,1,2} + T_{U,a,2}$, where

$$T_{U,a,1,1} = \sum_{(j_1,j_2) \in J_A^c} K(0, j_1, j_2), \quad T_{U,a,1,2} = \sum_{(j_1,j_2) \in J_A^c} \sum_{c=1}^{a} K(c, j_1, j_2), \tag{B.9}$$
$$T_{U,a,2} = \sum_{(j_1,j_2) \in J_A} \sum_{c=0}^{a} K(c, j_1, j_2).$$

Proof. Similarly to Section B.1.1, we first derive the variances and the covariances of the U-statistics, and then prove the asymptotic joint normality of the U-statistics. Particularly, the next Lemma B.1.11 derives the asymptotic form of $\mathrm{var}\{\mathcal{U}(a)\}$, and additionally shows that among the three terms in (B.9), $T_{U,a,1,1}$ is the leading one.

Lemma B.1.11. *Under the conditions of Theorem 3.2.5,* $\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\} \simeq$

$\mathrm{var}(T_{U,a,1,1})$, *where*

$$\mathrm{var}(T_{U,a,1,1}) \simeq 2a!\kappa_1^a n^{-a} \sum_{1\leq j_1\neq j_2\leq p} \sigma_{j_1,j_1}^a \sigma_{j_2,j_2}^a,$$

which is $\Theta(p^2 n^{-a})$*. Moreover,* $\mathrm{var}(T_{U,a,1,2}) = o(p^2 n^{-a})$, $\mathrm{var}(T_{U,a,2}) = o(p^2 n^{-a})$ *and* $\{\mathcal{U}(a) - T_{U,a,1,1}\}/\sigma(a) \xrightarrow{P} 0$.

Proof. See Section B.5.12 on Page 357. □

The following Lemma B.1.12 shows that the covariance between two different U-statistics asymptotically converges to 0.

Lemma B.1.12. *Under the conditions of Theorem 3.2.5, for two integers* $a \neq b$, $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} \to 0$.

Proof. See Section B.5.13 on Page 370. □

To finish the proof, it remains to show that the joint distribution of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ is asymptotically normal. By the Cramér-Wold theorem, it is equivalent to prove any fixed linear combination of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ converges to a normal distribution. By Lemma B.1.11, $\{\mathcal{U}(a) - T_{U,a,1,1}\}/\sigma(a) \xrightarrow{P} 0$. Thus by the Slutsky's theorem, it suffices to prove that any fixed linear combination of $[T_{U,a_1,1,1}/\sigma(a_1), \ldots, T_{U,a_m,1,1}/\sigma(a_m)]^\intercal$ converges to a normal distribution. Similarly to Section B.1.1, we redefine Z_n as below with $\sum_{r=1}^m t_r^2 = 1$, and prove that

$$Z_n := \sum_{r=1}^m t_r T_{U,a_r,1,1}/\sigma(a_r) \xrightarrow{D} \mathcal{N}(0,1). \tag{B.10}$$

We next prove (B.10) by the martingale central limit theorem, similarly to Section B.1.1. In particular, we define $\mathrm{E}_k(\cdot)$ in the same way as in Section B.1.1, and still define $D_{n,k} = (\mathrm{E}_k - \mathrm{E}_{k-1})Z_n$ and $\pi_{n,k}^2 = \mathrm{E}_{k-1}(D_{n,k}^2)$. It follows that

$D_{n,k} = \sum_{r=1}^m t_r A_{n,k,a_r}$ and $\pi_{n,k}^2 = \sum_{1\leq r_1,r_2\leq m} t_{r_1}t_{r_2}\mathrm{E}_{k-1}(A_{n,k,a_{r_1}}A_{n,k,a_{r_2}})$, where we redefine $A_{n,k,a_r} = (\mathrm{E}_k - \mathrm{E}_{k-1})\{T_{U,a_r,1,1}/\sigma(a_r)\}$. Note that $\sigma_{j_1,j_2} = 0$ when $(j_1,j_2) \in J_A^c$, and $T_{U,a,1,1}$ is a summation over $(j_1,j_2) \in J_A^c$. Thus the proof of Lemma B.1.4 in Section B.5.5 applies similarly, and we obtain the explicit form of $A_{n,k,a}$. Specifically, for each finite integer a, when $k < a$, $A_{n,k,a} = 0$; when $k \geq a$,

$$A_{n,k,a} = \frac{a}{\sigma(a)P_a^n} \sum_{1\leq i_1\neq\cdots\neq i_{a-1}\leq k-1} \sum_{(j_1,j_2)\in J_A^c} (x_{k,j_1}x_{k,j_2}) \prod_{t=1}^{a-1}(x_{i_t,j_1}x_{i_t,j_2}).$$

With the form of $A_{n,k,a}$, we can obtain the explicit forms of $D_{n,k}$ and $\pi_{n,k}^2$. Then we can prove the following two Lemmas B.1.13 and B.1.14, which suggests that (B.10) holds.

Lemma B.1.13. *Under the conditions of Theorem 3.2.5,* $\mathrm{var}(\sum_{k=1}^n \pi_{n,k}^2) \to 0$.

Proof. *See Section B.5.14 on Page 370.* ☐

Lemma B.1.14. *Under the conditions of Theorem 3.2.5 ,* $\sum_{k=1}^n \mathrm{E}(D_{n,k}^4) \to 0$.

Proof. *See Section B.5.15 on Page 380.* ☐

By Lemmas B.1.13 and B.1.14, (B.10) holds and thus Theorem 3.2.5 is proved.

B.1.5 Proof of Proposition 3.2.1

Consider the setting when n, p and $|J_A|$ are given and the value of M is fixed as $\Theta(1)$. We next examine ρ_a in (3.14) as a function of integer a in the following two cases.

(i) $|J_A| > Mp$ When $Mp/|J_A| < 1$, both $(Mp/|J_A|)^{1/a}$ and $(a!)^{1/(2a)}$ are increasing functions of integer a. Thus ρ_a is an increasing function of a. Since $a \in \mathbb{Z}^+$, ρ_a reaches the minimum value at $a = 1$.

(ii) $|J_A| \le Mp$ Define $\tilde{M} = Mp/|J_A|$, and $f(a) = (a!)^{1/(2a)}(\tilde{M})^{1/a}$. Note that ρ_a and $f(a)$ only differs by a constant. To find the minimum of ρ_a, it suffices to examine the minimum of $f(a)$.

In the following, we show that when $f(a)$ starts to not decrease at some value, it will strictly increase afterwards. Specifically, we prove that $f(a+2)/f(a+1) > 1$ if $f(a+1)/f(a) \ge 1$. Note that

$$
\begin{aligned}
\frac{f(a+1)}{f(a)} &= \frac{\{(a+1)!\}^{\frac{1}{2(a+1)}}(\tilde{M})^{\frac{1}{a+1}}}{(a!)^{\frac{1}{2a}}(\tilde{M})^{\frac{1}{a}}} \\
&= \left[\frac{\{(a+1)!\}^a \tilde{M}^{2a}}{(a!)^{a+1}\tilde{M}^{2(a+1)}}\right]^{\frac{1}{2a(a+1)}} = \{d(a) \times \tilde{M}^{-2}\}^{\frac{1}{2a(a+1)}},
\end{aligned}
$$

where $d(a) = (a+1)^a(a!)^{-1}$. It follows that $f(a+1)/f(a) > 1$ and $f(a+1)/f(a) = 1$ are equivalent to $d(a) > \tilde{M}^2$ and $d(a) = \tilde{M}^2$, respectively. We next show that $d(a)$ is a strictly increasing function of a. In particular,

$$
\frac{d(a+1)}{d(a)} = \frac{(a+2)^{a+1}a!}{(a+1)^a(a+1)!} = \left(\frac{a+2}{a+1}\right)^{a+1} > 1.
$$

Therefore we have $d(a+1) > \tilde{M}^2$ if $d(a) \ge \tilde{M}^2$, and equivalently this implies that $f(a+2)/f(a+1) > 1$ if $f(a+1)/f(a) \ge 1$.

Suppose a_0 is the first integer such that $d(a_0) \ge \tilde{M}^2$, i.e., for any integer $1 \le a < a_0$, $d(a) < \tilde{M}^2$. By the analysis above, we know $f(a)$ is decreasing when $a < a_0$, and $f(a)$ is strictly increasing when $a > a_0$. Thus a_0 achieves the minimum of $f(a)$, and a_0 increases as $\tilde{M}$ increases. Therefore the second part of proposition 3.2.1 is proved.

B.1.6 Proof of Proposition 3.2.2

Proof. Consider the simplified test statistic given in (3.16). We assume $\mathrm{E}(x_{i,j}) = 0$ and $\mathrm{var}(x_{i,j}^2) = 1$, $\forall j = 1, \ldots, p$ without loss of generality. It is then equivalent to examine $\mathcal{U}(\infty) = \max_{1\le j_1<j_2\le p} |\sum_{k=1}^n x_{k,j_1}x_{k,j_2}/n|$. We next prove (i) and (ii) of

Proposition 3.2.2 in the following Sections B.1.6.1 and B.1.6.2, respectively.

B.1.6.1 Proof of (i)

Under the alternative, we consider n i.i.d. observations $(x_{k,1}, x_{k,2})$, satisfying $\mathrm{E}(x_{k,1}x_{k,2}) = \rho$, for $k = 1, \ldots, n$. Then by Condition 3.2.2*, $\mathrm{var}(x_{k,1}x_{k,2}) = \mathrm{E}(x_{k,1}^2 x_{k,2}^2) - [\mathrm{E}(x_{k,1}x_{k,2})]^2 = \kappa_1(1 + 2\rho^2) - \rho^2$. The power of $\mathcal{U}(\infty)$ satisfies that

$$\begin{aligned} & P(|\,\mathcal{U}(\infty)\,| \geq t_p) \qquad\qquad \text{(B.11)} \\ = \;& P\Big(\max_{1\leq j_1<j_2\leq p}\Big|\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n\Big| \geq t_p\Big) \\ \geq \;& P\Big(\Big|\sum_{k=1}^{n} x_{k,1}x_{k,2}/n\Big| \geq t_p\Big) \\ = \;& P\left(\frac{\sum_{k=1}^{n}(x_{k,1}x_{k,2} - \rho)}{\sqrt{n}\sqrt{\mathrm{var}(x_{k,1}x_{k,2})}} \geq \frac{\sqrt{n}(t_p - \rho)}{\sqrt{\mathrm{var}(x_{k,1}x_{k,2})}}\right). \end{aligned}$$

We apply the central limit theorem on $x_{k,1}x_{k,2}$, $k = 1, \ldots, n$, and obtain

$$\frac{\sum_{k=1}^{n}(x_{k,1}x_{k,2} - \rho)}{\sqrt{n}\sqrt{\mathrm{var}(x_{k,1}x_{k,2})}} \xrightarrow{D} \mathcal{N}(0,1).$$

Suppose Z follows a standard Gaussian distribution. As $\log p \to \infty$, $\log p/n = o(1)$, and by Berry-Esseen Theorem, we have

$$\begin{aligned} \text{(B.11)} \;\geq\;& P\left(Z \geq \frac{\sqrt{n}(t_p - \rho)}{\sqrt{\mathrm{var}(x_{k1}x_{k2})}}\right) - \frac{C\mathrm{E}|x_{k1}x_{k2}|^3}{[\mathrm{var}(x_{k1}x_{k2})]^{\frac{3}{2}}\sqrt{n}} \\ \geq\;& P\left(Z \geq \frac{\sqrt{n}[n^{-1/2}\sqrt{4\log p} - \rho]}{\sqrt{\kappa_1(1+2\rho^2) - \rho^2}}\right) - \frac{C\sqrt{\mathrm{E}|x_{k1}|^6\mathrm{E}|x_{k2}|^6}}{[\mathrm{var}(x_{k1}x_{k2})]^{\frac{3}{2}}\sqrt{n}} \\ \geq\;& P(Z \geq C(2 - c_1)\sqrt{\log p}) - \frac{C}{\sqrt{n}} \\ \to\;& 1 + o(1), \end{aligned}$$

where the second inequality uses $t_p \le n^{-1/2}\sqrt{4\log p}$ when p is sufficiently large; the third inequality uses $\rho \ge c_1\sqrt{\log p/n}$; and the last step of convergence holds when $c_1 > 2$.

B.1.6.2 Proof of (ii)

Recall the notation J_A and J_A^c in Section B.1.4. Under the considered alternative, when $(j_1, j_2) \in J_A$, $\mathrm{E}(x_{k,j_1}x_{k,j_2}) = \rho$; and when $(j_3, j_4) \in J_A^c$, $\mathrm{E}(x_{k,j_3}x_{k,j_4}) = 0$. We have

$$
\begin{aligned}
P(|\mathcal{U}(\infty)| \ge t_p) \quad &\le \quad \sum_{1\le j_1<j_2\le p} P\Big(\Big|\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n\Big| \ge t_p\Big) \qquad\qquad \text{(B.12)}\\
&\le \quad \frac{1}{2}\sum_{(j_1,j_2)\in J_A} P\Big(\Big|\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n\Big| \ge t_p\Big)\\
&\quad\ +\frac{1}{2}\sum_{(j_3,j_4)\in J_A^c} P\Big(\Big|\sum_{k=1}^{n} x_{k,j_3}x_{k,j_4}/n\Big| \ge t_p\Big).
\end{aligned}
$$

Next we show that under the conditions of Proposition 3.2.2,

$$
\sum_{(j_1,j_2)\in J_A} P\Big(\Big|\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n\Big| \ge t_p\Big) \to 0, \qquad \text{(B.13)}
$$

and

$$
\frac{1}{2}\sum_{(j_3,j_4)\in J_A^c} P\Big(\Big|\sum_{k=1}^{n} x_{k,j_3}x_{k,j_4}/n\Big| \ge t_p\Big) \le \log(1-\alpha)^{-1}. \qquad \text{(B.14)}
$$

Proof of (B.13)**.** To prove (B.13), we will next derive an upper bound of the probability $P(|\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n| \ge t_p)$ for each $(j_1, j_2) \in J_A$ by Lemma 6.8 in Cai and Jiang (2011). In the following, we consider a fixed index pair (j_1, j_2), and for easy presentation, we write $m_0 = \sqrt{\mathrm{var}(x_{k,j_1}x_{k,j_2})}$ and $\xi_k = (x_{k,j_1}x_{k,j_2} - \rho)/m_0$. When $(j_1, j_2) \in J_A$, we have $\mathrm{E}(\xi_k) = 0$, $\mathrm{var}(\xi_k) = 1$, and by Condition 3.2.2*, $m_0^2 = \kappa_1(1 + 2\rho^2) - \rho^2$. It follows that

$$P\Big(\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n \geq t_p\Big) = P\Big(\frac{\sum_{k=1}^n \xi_k}{\sqrt{n\log p}} \geq y_n\Big),$$

where $y_n = \sqrt{n/\log p}\, m_0^{-1}(t_p - \rho)$. We next show that y_n and $\xi_k, k = 1, \ldots, n$ satisfy the conditions of Lemma 6.8 in Cai and Jiang (2011). First note that $y_n \to y = (2-c_2)m_0^{-1}$, and $y > 0$ as $c_2 < 2$. We then show that $\mathrm{E}\{\exp(\tilde{t}_0|\xi_k|^{\vartheta})\} < \infty$ for some $\tilde{t}_0 > 0$ and $0 < \vartheta \leq 1$. In particular, given ς and t_0 in Proposition 3.2.2, we take $\vartheta = \varsigma/2 \in (0, 1]$ and $\tilde{t}_0 = t_0(2m_0)^{\vartheta}/2 > 0$. By Lemma B.5.4,

$$\begin{aligned}
&|x_{k,j_1}x_{k,j_2} - \rho|^{\vartheta} \leq (|x_{k,j_1}x_{k,j_2}| + |\rho|)^{\vartheta} \leq |x_{k,j_1}x_{k,j_2}|^{\vartheta} + |\rho|^{\vartheta} \\
&\leq \Big(\frac{x_{k,j_1}^2 + x_{k,j_2}^2}{2}\Big)^{\vartheta} + |\rho|^{\vartheta} \leq \frac{1}{2^{\vartheta}}(|x_{k,j_1}|^{2\vartheta} + |x_{k,j_2}|^{2\vartheta}) + |\rho|^{\vartheta}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{E}\exp(\tilde{t}_0|\xi_k|^{\vartheta}) \leq& \mathrm{E}\exp\Big[\frac{\tilde{t}_0}{(2m_0)^{\vartheta}}(|x_{k,j_1}|^{2\vartheta} + |x_{k,j_2}|^{2\vartheta}) + \frac{\tilde{t}_0}{m_0^{\vartheta}}|\rho|^{\vartheta}\Big] \qquad \text{(B.15)} \\
=& \mathrm{E}[\exp(2^{-1}t_0|x_{k,j_1}|^{\varsigma}) \times \exp(2^{-1}t_0|x_{k,j_2}|^{\varsigma})] \times \exp(t_0 2^{\vartheta-1}|\rho|^{\vartheta}) \\
\leq& \sqrt{\mathrm{E}[\exp(t_0|x_{k,j_1}|^{\varsigma})] \times \mathrm{E}[\exp(t_0|x_{k,j_2}|^{\varsigma})]} \times \exp(t_0 2^{\vartheta-1}|\rho|^{\vartheta}),
\end{aligned}$$

where the last inequality follows from the Hölder's inequality. By the conditions in Proposition 3.2.2, we know $\max_{(j_1,j_2)\in J_A} \mathrm{E}(t_0|x_{k,j_1}|^{\varsigma}) \times \mathrm{E}(t_0|x_{k,j_2}|^{\varsigma}) < \infty$ and $\rho \leq c_2\sqrt{\log p/n} = o(1)$. Therefore, (B.15) $< \infty$. In summary, y_n and $\xi_k, k = 1, \ldots, n$ satisfy the conditions of Lemma 6.8 in Cai and Jiang (2011).

By Lemma 6.8 in Cai and Jiang (2011), as $\log p = o(n^{\beta})$ and $\beta = \vartheta/(2+\vartheta) = \varsigma/(4+\varsigma)$,

$$P\Big(\frac{\sum_{k=1}^n \xi_k}{\sqrt{n\log p}} \geq y_n\Big) \simeq \frac{p^{-y_n^2/2}(\log p)^{-1/2}}{\sqrt{2\pi}y}. \qquad \text{(B.16)}$$

Let $z_0 = -\log(8\pi) - 2\log\log(1-\alpha)^{-1}$, then we can write $t_p = n^{-1/2}\{4\log p - \log\log p + z_0\}^{1/2}$ and

$$\begin{aligned}
y_n^2 &= \frac{n}{\log p}(t_p - \rho)^2 \times \frac{1}{\mathrm{var}(x_{k,j_1}x_{k,j_2})} \\
&= \frac{n}{\log p}(t_p^2 - 2\rho t_p + \rho^2) \times \frac{1}{\mathrm{var}(x_{k,j_1}x_{k,j_2})} \\
&\geq \frac{1}{\mathrm{var}(x_{k,j_1}x_{k,j_2})} \times \Big\{\frac{1}{\log p}\Big(4\log p - \log\log p + z_0\Big) \\
&\quad -\frac{2c_2\sqrt{\log p}\sqrt{4\log p - \log\log p + z_0}}{\log p} + \frac{c_2^2 \log p}{\log p}\Big\},
\end{aligned}$$

where the last inequality holds when $\rho \leq c_2\sqrt{\log p/n}$ and $c_2 < 2$. Then

$$\begin{aligned}
& p^{-y_n^2/2} \\
&= \exp(-(\log p)y_n^2/2) \\
&\leq \exp\Bigg\{-\frac{1}{\mathrm{var}(x_{k,j_1}x_{k,j_2})}\Big[\frac{1}{2}\Big(4\log p - \log\log p - \log(8\pi) - 2\log\log(1-\alpha)^{-1}\Big) \\
&\qquad - c_2\sqrt{\log p}\sqrt{4\log p - \log\log p + z_0} + \frac{c_2^2\log p}{2}\Big]\Bigg\} \\
&= \Big\{p^{-2}\sqrt{\log p} \times \sqrt{8\pi}\log(1-\alpha)^{-1} \times p^{-\frac{c_2^2}{2}} \\
&\qquad \times \exp\Big(c_2\sqrt{\log p}\sqrt{4\log p - \log\log p + z_0}\Big)\Big\}^{1/\{\mathrm{var}(x_{k,j_1}x_{k,j_2})\}}.
\end{aligned}$$

By Condition 3.2.2*, $\mathrm{var}(x_{k,j_1}x_{k,j_2}) = \kappa_1 + (2\kappa_1 - 1)\rho^2$, and as $\rho = o(1)$, there exists a constant $m > 0$ such that $\mathrm{var}(x_{k,j_1}x_{k,j_2}) \leq \kappa_1 + m$. Thus

$$\begin{aligned}
& p^{-y_n^2/2}(\log p)^{-1/2} \\
\leq\ & (\log p)^{-1/2}\Big\{p^{-2}\sqrt{\log p} \times \sqrt{8\pi}(\log(1-\alpha)^{-1})p^{-\frac{c_2^2}{2}} \\
& \times \exp\Big(c_2\sqrt{\log p}\sqrt{4\log p - \log\log p + z_0}\Big)\Big\}^{1/\{\mathrm{var}(x_{k1}x_{k2})\}} \\
\leq\ & (\log p)^{-1/2}\Big[\sqrt{8\pi}\log(1-\alpha)^{-1}\sqrt{\log p} \times p^{-2-\frac{c_2^2}{2}+2c_2}\Big]^{1/(\kappa_1+m)}.
\end{aligned}$$

Recall that $y = (2 - c_2)[\mathrm{var}(x_{k,j_1}x_{k,j_2})]^{-1/2}$. Then by (B.16),

$$\begin{aligned}
&\frac{1}{2}\sum_{(j_1,j_2)\in J_A} P\Big(\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}/n \geq t_p\Big) \qquad \text{(B.17)}\\
&= \frac{1}{2}\sum_{(j_1,j_2)\in J_A} P\Big(\frac{\sum_{k=1}^{n}\xi_k}{\sqrt{n\log p}} \geq y_n\Big)\\
&\leq \frac{|J_A|}{2}\frac{(\log p)^{-1/2}}{y\sqrt{2\pi}}\Big(\sqrt{8\pi}\log(1-\alpha)^{-1}\sqrt{\log p}\times p^{-2-\frac{c_2^2}{2}+2c_2}\Big)^{\frac{1}{\kappa_1+m}}\\
&= C_\alpha \exp\left(2\log\Big[p^{-\frac{(1-c_2+c_2^2/4)}{(\kappa_1+m)}}\Big\{\sqrt{|J_A|}(\log p)^{\frac{1}{4(\kappa_1+m)}-\frac{1}{4}}\Big\}\Big]\right),
\end{aligned}$$

where $C_\alpha = \frac{1}{2y\sqrt{2\pi}}[\sqrt{8\pi}\log(1-\alpha)^{-1}]^{1/(\kappa_1+m)}$. Thus, (B.17) $\to 0$ when

$$p^{-\frac{(1-c_2/2)^2}{\kappa_1+m}}\sqrt{|J_A|}(\log p)^{\frac{1}{4(\kappa_1+m)}-\frac{1}{4}} \to 0.$$

Similarly, we have

$$\begin{aligned}
&\sum_{(j_1,j_2)\in J_A} P\Big(\frac{\sum_{k=1}^{n} x_{k,j_1}x_{k,j_2}}{n} \leq -t_p\Big) \qquad \text{(B.18)}\\
&= \sum_{(j_1,j_2)\in J_A} P\Big(\frac{\sum_{k=1}^{n}(-x_{k,j_1}x_{k,j_2}+\rho)}{n\sqrt{\mathrm{var}(x_{k,j_1}x_{k,j_2})}} \geq \frac{t_p+\rho}{\sqrt{\mathrm{var}(x_{k,j_1}x_{k,j_2})}}\Big),
\end{aligned}$$

and (B.18) $\to 0$ following the similar arguments as above. In summary, (B.13) holds when $J_A = o(1)p^{\frac{2(1-c_2/2)^2}{\kappa_1+m}}(\log p)^{\frac{1}{2}-\frac{1}{2(\kappa_1+m)}}$ for some $m > 0$.

Proof of (B.14)**.** Similarly to Section B.1.6.2, we derive an upper bound of the probability $P(\sum_{k=1}^{n} x_{k,j_3}x_{k,j_4}/n \geq t_p)$ for each $(j_3, j_4) \in J_A^c$ by Lemma 6.8 in Cai and Jiang (2011). In the following, we consider a fixed index pair (j_3, j_4); and for easy presentation, we write $\tilde{\xi}_k = x_{k,j_3}x_{k,j_4}/\sqrt{\kappa_1}$, $k = 1, \ldots, n$. When $(j_3, j_4) \in J_A^c$, $\mathrm{E}(x_{k,j_3}x_{k,j_4}) = 0$ and $\mathrm{var}(x_{k,j_3}x_{k,j_4}) = \mathrm{E}\{(x_{k,j_3}x_{k,j_4})^2\} = \kappa_1$, then we have $\mathrm{E}(\tilde{\xi}_k) = 0$

and $\text{var}(\tilde{\xi}_k) = 1$. To prove (B.14), we write

$$P\Big(\sum_{k=1}^{n} x_{k,j_3}x_{k,j_4}/n \geq t_p\Big) = P\Big(\frac{\sum_{k=1}^{n}\tilde{\xi}_k}{\sqrt{n\log p}} \geq \tilde{y}_n\Big),$$

where $\tilde{y}_n = \sqrt{n/\log p} \times t_p/\sqrt{\kappa_1} \to \tilde{y} = 2/\sqrt{\kappa_1}$. Similarly to Section B.1.6.2, we know $\tilde{y}_n$ and $\tilde{\xi}_k$, $k = 1, \ldots, n$ also satisfy the conditions of Lemma 6.8 in Cai and Jiang (2011). Thus by Lemma 6.8 in Cai and Jiang (2011), for $z_0 = -\log(8\pi) - 2\log\log(1-\alpha)^{-1}$ and $t_p = n^{-1/2}\sqrt{4\log p - \log\log p + z_0}$,

$$\begin{aligned}
& P\Big(\frac{\sum_{k=1}^{n}\tilde{\xi}_k}{\sqrt{n\log p}} \geq \tilde{y}_n\Big) \\
\simeq \quad & \frac{p^{-\tilde{y}_n^2/2}(\log p)^{-1/2}}{\sqrt{2\pi}\tilde{y}} \\
= \quad & p^{-2/\kappa_1}(\log p)^{1/(2\kappa_1)-1/2}\frac{\exp(-z_0/(2\kappa_1))}{\sqrt{2\pi}\tilde{y}} \\
\leq \quad & (8\pi)^{1/(2\kappa_1)}\frac{\sqrt{\kappa_1}}{2\sqrt{2\pi}}p^{-2/\kappa_1}(\log p)^{1/(2\kappa_1)-1/2}\{\log(1-\alpha)^{-1}\}^{1/\kappa_1}.
\end{aligned}$$

Then for $\kappa_1 \leq 1$ and a small $\alpha > 0$,

$$\begin{aligned}
& \frac{1}{2}\sum_{(j_1,j_2)\in J_A^c} P\Big(\sum_{k=1}^{n} x_{k,j_3}x_{k,j_4}/n \geq t_p\Big) \\
\leq \quad & \frac{1}{2}\frac{p(p-1)-|J_A|}{p^{2/\kappa_1}(\log p)^{-1/(2\kappa_1)+1/2}}(8\pi)^{1/(2\kappa_1)}\frac{\sqrt{\kappa_1}}{2\sqrt{2\pi}}\{\log(1-\alpha)^{-1}\}^{1/\kappa_1},
\end{aligned} \tag{B.19}$$

which attains the maximum order at $\kappa_1 = 1$, when $\kappa_1 \leq 1$ and $n, p \to \infty$. Therefore asymptotically, (B.19) $\leq 2^{-1}\log(1-\alpha)^{-1}$. By similar arguments, we know when $n, p \to \infty$,

$$\frac{1}{2}\sum_{(j_3,j_4)\in J_A^c} P\Big(\sum_{k=1}^{n} x_{k,j_3}x_{k,j_4}/n \leq -t_p\Big) \leq \frac{1}{2}\log(1-\alpha)^{-1}.$$

In summary, we have (B.14) holds.

Combining (B.13) and (B.14), we obtain (B.12) $\leq \log(1-\alpha)^{-1}$. $\square$

B.2 Proofs of Theoretical Results in Section 3.3

B.2.1 Proof of Theorems 3.3.1 and 3.3.2

Under H_0, for $\mathcal{U}(a)$ in (3.19), we assume without loss of generality that $\boldsymbol{\mu}_0 = \mathbf{0}$, and then write $\mathcal{U}(a) = \sum_{j=1}^{p} (P_a^n)^{-1} \sum_{1\leq i_1 \neq \cdots \neq i_a \leq n} \prod_{k=1}^{a} x_{i_k,j}$.

We start with the proof of Theorem 3.3.1. Similarly to Section B.1.1, we first derive the variances and the covariances of the U-statistics; and then prove the asymptotic joint normality of the U-statistics. In particular, for $\mathrm{var}\{\mathcal{U}(a)\}$ in Theorem 3.3.1, as $\mathrm{E}\{\mathcal{U}(a)\} = 0$ under H_0,

$$\mathrm{var}\{\mathcal{U}(a)\} = \mathrm{E}\{\mathcal{U}^2(a)\} = (P_a^n)^{-2} \sum_{\substack{1\leq j_1 \leq p,\\ 1\leq j_2 \leq p}} \sum_{\substack{1\leq i_1 \neq \cdots \neq i_a \leq n,\\ 1\leq \tilde{i}_1 \neq \cdots \neq \tilde{i}_a \leq n}} \mathrm{E}\Big(\prod_{k=1}^{a} x_{i_k,j_1} x_{\tilde{i}_k,j_2} \Big).$$

Note that $\mathrm{E}(\prod_{k=1}^{a} x_{i_k,j_1} \times x_{\tilde{i}_k,j_2}) = 0$ when $\{i_1,\ldots,i_a\} \neq \{\tilde{i}_1,\ldots,\tilde{i}_a\}$. Moreover, $\mathrm{E}(\prod_{k=1}^{a} x_{i_k,j_1} x_{\tilde{i}_k,j_2}) = \sigma_{j_1,j_2}^a$ when $\{i_1,\ldots,i_a\} = \{\tilde{i}_1,\ldots,\tilde{i}_a\}$. Then

$$\mathrm{var}\{\mathcal{U}(a)\} = (P_a^n)^{-1} \sum_{1\leq j_1,j_2 \leq p} a! \sigma_{j_1,j_2}^a. \tag{B.20}$$

By Condition 3.3.1, $\sum_{1\leq j_1,j_2\leq p} \sigma_{j_1,j_2}^a = \Theta(p)$. Thus $\mathrm{var}\{\mathcal{U}(a)\} = \Theta(pn^{-a})$.

Second, we show that $\mathrm{cov}\{\mathcal{U}(a),\mathcal{U}(b)\} = 0$. Note that under H_0, $\mathrm{cov}\{\mathcal{U}(a),\mathcal{U}(b)\} = \mathrm{E}\{\mathcal{U}(a)\mathcal{U}(b)\}$ and

$$\mathrm{E}\{\mathcal{U}(a)\mathcal{U}(b)\} = (P_a^n P_b^n)^{-1} \sum_{\substack{1\leq j_1 \leq p,\\ 1\leq j_2 \leq p}} \sum_{\substack{1\leq i_1 \neq \cdots \neq i_a \leq n,\\ 1\leq \tilde{i}_1 \neq \cdots \neq \tilde{i}_a \leq n}} \mathrm{E}\Big(\prod_{k=1}^{a} x_{i_k,j_1} \prod_{t=1}^{b} x_{\tilde{i}_t,j_2} \Big).$$

Since $a \neq b$, $\{i_1,\ldots,i_a\} \neq \{\tilde{i}_1,\ldots,\tilde{i}_b\}$. Suppose there exists an index $i \in \{i_1,\ldots,i_a\}$

and $i \notin \{\tilde{i}_1, \ldots, \tilde{i}_b\}$. Then under H_0,

$$\mathrm{E}\Big(\prod_{k=1}^{a} x_{i_k,j_1} \prod_{t=1}^{b} x_{\tilde{i}_t,j_2}\Big) = \mathrm{E}(x_{i,j})\mathrm{E}(\text{all the remaining terms}) = 0.$$

Therefore, $\mathrm{E}\{\mathcal{U}(a)\mathcal{U}(b)\} = 0$.

In summary, the covariance matrix of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ asymptotically converges to I_m. To finish the proof of Theorem 3.3.1, it remains to show that the joint limiting distribution of the U-statistics is normal. By the Cramér-Wold theorem, it is sufficient to prove that any fixed linear combination of these U-statistics converges to a normal distribution. Similarly to Section B.1.1, we use the martingale central limit theorem (Billingsley, 1995, p.476). Specifically, we redefine Z_n as below with $\sum_{r=1}^{m} t_r^2 = 1$, and prove that

$$Z_n := \sum_{r=1}^{m} t_r \mathcal{U}(a_r)/\sigma(a_r) \xrightarrow{D} \mathcal{N}(0,1). \tag{B.21}$$

With the redefined Z_n, we define $\mathrm{E}_k(\cdot)$ in the same way as in Section B.1.1, and still define $D_{n,k} = (\mathrm{E}_k - \mathrm{E}_{k-1})Z_n$ and $\pi_{n,k}^2 = \mathrm{E}_{k-1}(D_{n,k}^2)$. Similarly to Section B.1.1, we have $D_{n,k} = (\mathrm{E}_k - \mathrm{E}_{k-1})Z_n = \sum_{r=1}^{m} t_r A_{n,k,a_r}$, where we redefine $A_{n,k,a_r} = (\mathrm{E}_k - \mathrm{E}_{k-1})\{\mathcal{U}(a_r)/\sigma(a_r)\}$. In addition, similarly to Lemma B.1.4, we obtain that when $k < a_r$, $A_{n,k,a_r} = 0$; and when $k \geq a_r$,

$$A_{n,k,a_r} = \frac{a_r}{\sigma(a_r) P_{a_r}^n} \sum_{j=1}^{p} \sum_{1 \leq i_1 \neq \cdots \neq i_{a_r-1} \leq k-1} x_{k,j} \times \prod_{t=1}^{a_r-1} x_{i_t,j}.$$

Given the form of A_{n,k,a_r}, we can obtain the forms of $D_{n,k}$ and $\pi_{n,k}^2$. To prove (B.10), by the martingale central limit theorem, it suffices to prove the following Lemma B.2.1.

Lemma B.2.1. *Under the conditions of Theorem 3.3.1,* $\mathrm{var}(\sum_{k=1}^{n} \pi_{n,k}^2) \to 0$ *and* $\sum_{k=1}^{n} \mathrm{E}(D_{n,k}^4) \to 0$.

Proof. See Section B.5.16 on Page 383. $\square$

With Lemma B.2.1, the asymptotic joint normality in Theorem 3.3.1 is obtained by the martingale central limit theorem. For Theorem 3.3.2, the limiting distribution of $\mathcal{U}(\infty)$ follows from Cai et al. (2014). In addition, the asymptotic independence between $\mathcal{U}(a)/\sigma(a)$ and $n\mathcal{U}(\infty) - \tau_p$ can be obtained similarly as the proof of Theorem 3.3.4. We defer the details to Section B.2.3.

B.2.2 Proof of Theorem 3.3.3

By the following Proposition B.2.1, we assume that under H_0, $\boldsymbol{\mu} = \boldsymbol{\nu} = \mathbf{0}$, without loss of generality.

Proposition B.2.1. *$\mathcal{U}(a)$ constructed in (3.20) and (3.21) are location invariant; that is, for any vector $\boldsymbol{\Delta} \in \mathbb{R}^p$, the U-statistic constructed based on the transformed data $\{\mathbf{x}_i + \Delta : i = 1, \ldots, n_x\}$ and $\{\mathbf{y}_i + \Delta : i = 1, \ldots, n_y\}$ is still $\mathcal{U}(a)$.*

Proposition B.2.1 can be obtained straightforwardly from the definitions $\mathcal{U}(a) = \sum_{j=1}^{p} (P_a^{n_x} P_a^{n_y})^{-1} \sum_{\substack{1 \le k_1 \neq \ldots \neq k_a \le n_x; \\ 1 \le s_1 \neq \ldots \neq s_a \le n_y}} \prod_{t=1}^{a} (x_{k_t,j} - y_{s_t,j})$ and $\mathcal{U}(\infty) = \max_{1 \le j \le p} \sigma_{j,j}^{-1} \times (\bar{x}_j - \bar{y}_j)^2$ in (3.21). The proof is thus skipped.

The following proof proceeds by deriving the variances, covariances and asymptotic joint normality of the U-statistics. Particularly, the next Lemma B.2.2 derives the asymptotic form of $\sigma^2(a)$ in Theorem 3.3.3.

Lemma B.2.2. *Under the conditions of Theorem 3.3.3,*

$$\mathrm{var}\{\mathcal{U}(a)\} \simeq \sum_{1 \le j_1, j_2 \le p} a! \Big(\frac{\sigma_{x,j_1,j_2}}{n_x} + \frac{\sigma_{y,j_1,j_2}}{n_y} \Big)^a = \Theta(pn^{-a}).$$

When $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2} = \sigma_{j_1,j_2}$, we have $\mathrm{var}[\mathcal{U}(a)] \simeq \sum_{j_1,j_2=1}^{p} a!(n_x + n_y)^a \sigma_{j_1,j_2}^a / (n_x n_y)^a$.

Proof. See Section B.5.17 on Page 388. $\square$

In addition, the following Lemma B.2.3 shows that different $\mathcal{U}(a)$'s of finite a are uncorrelated.

Lemma B.2.3. *Under the conditions of Theorem 3.3.3, for finite integers $a \neq b$,* $\text{cov}\{\mathcal{U}(a), \mathcal{U}(b)\} = 0$.

Proof. *See Section B.5.18 on Page 390.* □

We then know $\text{cov}\{\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)\} = I_m$ by Lemmas B.2.2 and B.2.3. The next Lemma B.2.4 further proves the asymptotic joint normality of the U-statistics.

Lemma B.2.4. *Under the conditions of Theorem 3.3.3, for finite integers $a_1, \ldots, a_m$,* $\{\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)\} \xrightarrow{D} \mathcal{N}(0, I_m)$.

Proof. *See Section B.5.19 on Page 390.* □

Combining Lemmas B.2.2–B.2.4, we finish the proof of Theorem 3.3.3.

B.2.3 Proof of Theorem 3.3.4

For $\mathcal{U}(\infty)$ in (3.21), the limiting distribution of $\mathcal{U}(\infty)$ is established in Cai et al. (2014) and Xu et al. (2016). We next prove the asymptotic independence between $\mathcal{U}(\infty)$ and $\mathcal{U}(a)$ by a similar argument to that in Hsing (1995), see also Xu et al. (2016). In this proof, we reserve the notation P for the probability measure on which $x_{i,j}$ and $y_{i,j}$ are defined, and the expectation with respect to P is denoted as E. Define $\tilde{\mathcal{U}}_c(a)/\sigma(a)$ on the conditional probability measure $\tilde{P}$, given the event $n_x n_y \mathcal{U}(\infty)/(n_x + n_y) - \tau_p \leq u$ such that

$$\tilde{P}\Big\{\tilde{\mathcal{U}}_c(a)/\sigma(a) \leq u'\Big\} = P\Big\{\mathcal{U}(a)/\sigma(a) \leq u' \;\Big|\; \frac{n_x n_y}{n_x + n_y}\mathcal{U}(\infty) \leq \tau_p + u\Big\}.$$

The expectation with respect to $\tilde{P}$ is denoted by $\tilde{\text{E}}$. To show the asymptotic independence, it is sufficient to prove the following Lemma B.2.5.

Lemma B.2.5. *Under the conditions of Theorem 3.3.4, $\tilde{\mathcal{U}}_c(a)/\sigma(a) \xrightarrow{D} \mathcal{N}(0,1)$ on the conditional measure $\tilde{P}$.*

Proof. *See Section B.5.20 on Page 396.* $\square$

B.2.4 Proof of Theorem 3.3.5

By Proposition B.2.1, we assume $\mathrm{E}(\mathbf{y}) = \boldsymbol{\nu} = \mathbf{0}$, without loss of generality. Then under the considered alternative $\mathcal{E}_A$, $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu} = \{\mu_j = \rho : j = 1, \dots, k_0; \mu_j = 0 : j = k_0 + 1, \dots, p\}$. Define $\varphi_{j_1,j_2} = \sigma_{j_1,j_2} + \mu_{j_1}\mu_{j_2}$. We have $\mathrm{E}(x_{i,j_1}x_{i,j_2}) = \varphi_{j_1,j_2}$, and under $\boldsymbol{\nu} = \mathbf{0}$, $\mathrm{E}(y_{i,j_1}y_{i,j_2}) = \sigma_{j_1,j_2}$.

Similarly to the proof of Theorem 3.2.5 in Section B.1.4, we decompose $\mathcal{U}(a) = T_{a,1} + T_{a,2}$, where

$$T_{a,1} = \sum_{j=1}^{k_0}\sum_{c=0}^{a} \sum_{\substack{1\le k_1\neq\cdots\neq k_c\le n_x,\\ 1\le s_1\neq\cdots\neq s_{a-c}\le n_y}} G(a,c)\prod_{t=1}^{c} x_{k_t,j} \prod_{m=1}^{a-c} y_{s_m,j}, \tag{B.22}$$

$$T_{a,2} = \sum_{j=k_0+1}^{p}\sum_{c=0}^{a} \sum_{\substack{1\le k_1\neq\cdots\neq k_c\le n_x,\\ 1\le s_1\neq\cdots\neq s_{a-c}\le n_y}} G(a,c)\prod_{t=1}^{c} x_{k_t,j} \prod_{m=1}^{a-c} y_{s_m,j},$$

with $G(a,c) = (-1)^{a-c}\binom{a}{c}(P_c^{n_x}P_{a-c}^{n_y})^{-1}$. Then $\mathrm{E}(T_{a,1}) = \sum_{j=1}^{k_0}(\mu_j - \nu_j)^a = k_0\rho^a$ and $\mathrm{E}(T_{a,2}) = \sum_{j=k_0+1}^{p}(\mu_j - \nu_j)^a = 0$.

To prove Theorem 3.3.5, we derive the variances, covariances, and asymptotic joint normality of the U-statistics. Particularly, the next Lemma B.2.6 gives the asymptotic form of $\sigma^2(a) = \mathrm{var}\{\mathcal{U}(a)\}$, and shows that $T_{a,2}$ is the leading component.

Lemma B.2.6. *Under the conditions of Theorem 3.3.5,*

$$\mathrm{var}\{\mathcal{U}(a)\} \simeq \sum_{k_0+1\le j_1,j_2\le p} a!\Big(\frac{\sigma_{x,j_1,j_2}}{n_x} + \frac{\sigma_{y,j_1,j_2}}{n_y}\Big)^a. \tag{B.23}$$

$\mathrm{var}(T_{a,2}) = \Theta(pn^{-a})$ *and* $\mathrm{var}(T_{a,1}) = o(1)\mathrm{var}(T_{a,2})$. *Then* $\{T_{a,1} - \mathrm{E}(T_{a,1})\}/\sigma(a) \xrightarrow{P} 0.$

Proof. See Section B.5.21 on Page 398. $\square$

In addition, the following Lemma B.2.7 shows that the covariance between two U-statistics asymptotically converges to 0.

Lemma B.2.7. *Under the conditions of Theorem 3.3.5, for two finite integers $a \neq b$, $\{\sigma(a)\sigma(b)\}^{-1}\mathrm{cov}\{\mathcal{U}(a), \mathcal{U}(b)\} \to 0$.*

Proof. See Section B.5.22 on Page 401. $\square$

By the analysis above, we know the covariance matrix of $[\{\mathcal{U}(a_1) - \mathrm{E}[\mathcal{U}(a_1)]\}/\sigma(a_1), \ldots, \{\mathcal{U}(a_m) - \mathrm{E}[\mathcal{U}(a_m)]\}/\sigma(a_m)]^\intercal$ asymptotically converges to I_m. To prove Theorem 3.3.5, it remains to show that the joint limiting distribution of the U-statistics is normal. By the Cramér-Wold theorem, it is equivalent to prove that any fixed linear combination of these U-statistics converges to a normal distribution. By Lemma B.2.6 and the Slutsky's theorem, it suffices to show that any fixed linear combination of $[T_{a_1,2}/\sqrt{\mathrm{var}(T_{a_1,2})}, \ldots, T_{a_m,2}/\sqrt{\mathrm{var}(T_{a_m,2})}]^\intercal$ converges to a normal distribution for any finite m. Since $\mu_j = \nu_j$ for $j \in \{k_0+1, \ldots, p\}$, and each $T_{a_t,2}$ is a summation over $j \in \{k_0+1, \ldots, p\}$, we know the analysis under H_0 in Section B.2.2 can be applied to $T_{a_t,2}$ similarly. Given $k_0 = o(p)$, we know $[T_{a_1,2}/\sqrt{\mathrm{var}(T_{a_1,2})}, \ldots, T_{a_m,2}/\sqrt{\mathrm{var}(T_{a_m,2})}]^\intercal$ has the joint asymptotic normality. In summary, Theorem 3.3.5 is proved.

B.3 Proofs of Theoretical Results in Section 3.4

B.3.1 Proof of Theorem 3.4.1

Since $\mathcal{U}(a)$ is location invariant, we assume $\mathrm{E}(\mathbf{x}) = \mathbf{0}$ and $\mathrm{E}(\mathbf{y}) = \mathbf{0}$, without loss of generality, in this section. We decompose $\mathcal{U}(a) = \tilde{\mathcal{U}}(a) + \tilde{\mathcal{U}}^*(a)$, where we redefine

$$\tilde{\mathcal{U}}(a) = \sum_{1 \leq j_1, j_2 \leq p} \frac{1}{P_a^{n_x} P_a^{n_y}} \sum_{\substack{1 \leq i_1 \neq \ldots \neq i_a \leq n_x; \\ 1 \leq w_1 \neq \ldots \neq w_a \leq n_y}} \prod_{t=1}^{a} (x_{i_t,j_1} x_{i_t,j_2} - y_{w_t,j_1} y_{w_t,j_2}),$$

and $\tilde{\mathcal{U}}^*(a) = \mathcal{U}(a) - \tilde{\mathcal{U}}(a)$. To prove Theorem 3.4.1, we derive the variances, covariances, and asymptotic joint normality of the U-statistics. Particularly, the following Lemma B.3.1 derives the asymptotic form of $\mathrm{var}\{\mathcal{U}(a)\}$, and shows that $\tilde{\mathcal{U}}(a)$ is the leading term.

Lemma B.3.1. *Under the conditions of Theorem 3.4.1,* $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$, $\mathcal{U}^*(a)/\sigma(a) \xrightarrow{P} 0$, *and*

$$\mathrm{var}\{\mathcal{U}(a)\} \simeq \sum_{1\leq j_1,j_2,j_3,j_4\leq p} a!\left\{\frac{\Pi^x_{j_1,j_2,j_3,j_4} - \sigma_{j_1,j_2}\sigma_{j_3,j_4}}{n_x} + \frac{\Pi^y_{j_1,j_2,j_3,j_4} - \sigma_{j_1,j_2}\sigma_{j_3,j_4}}{n_y}\right\}^a.$$

In particular, under Condition 3.4.1, $\mathrm{var}\{\mathcal{U}(a)\} = \Theta(p^2 n^{-a})$; *under Condition 3.4.1*,* $\mathrm{var}\{\mathcal{U}(a)\} = \Theta\{n^{-a}\sum_{1\leq j_1,j_2,j_3,j_4\leq p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a\}$.

Proof. *See Section B.5.23 on Page 402.* □

Given Lemma B.3.1, the next Lemma B.3.2 shows that the covariance between two U-statistics asymptotically converges to 0.

Lemma B.3.2. *Under the conditions of Theorem 3.4.1, for finite integers $a \neq b$,* $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} \to 0$ *as $n, p \to \infty$.*

Proof. *See Section B.5.24 on Page 409.* □

To finish the proof, it remains to show that the joint distribution of $[\mathcal{U}(a_1)/\sigma(a_1), \ldots, \mathcal{U}(a_m)/\sigma(a_m)]^\intercal$ is asymptotically normal for different finite integers $a_1, \ldots, a_m$. By the Cramér-Wold theorem, it is equivalent to prove that any of their fixed linear combination converges to normal. In addition, by Lemma B.3.1 and Slutsky's theorem, it suffices to prove that any fixed linear combination of $[\tilde{\mathcal{U}}(a_1)/\sigma(a_1), \ldots, \tilde{\mathcal{U}}(a_m)/\sigma(a_m)]^\intercal$ converges to normal. Specifically, similarly to Section B.1.1, we redefine Z_n as below with $\sum_{r=1}^m t_r^2 = 1$, and prove that

$$Z_n := \sum_{r=1}^m t_r \tilde{\mathcal{U}}(a_r)/\sigma(a_r) \xrightarrow{D} \mathcal{N}(0,1). \tag{B.24}$$

We next prove (B.24) following the proof of Theorem 3.2.1 in Section B.1.1 and apply the martingale central limit theorem (Billingsley, 1995, p.476).

To construct a martingale difference, we write $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^\intercal$ and $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,p})^\intercal$; and define a new random vector

$$R_i = \mathbf{x}_i \quad \text{for } i = 1, 2, \ldots, n_x; \qquad R_{n_x+j} = \mathbf{y}_j \quad \text{for } j = 1, 2, \ldots, n_y.$$

We then define $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_k = \sigma\{R_1, \ldots, R_k\}$ for $k = 1, 2, \ldots, n_x + n_y$; and let $\mathrm{E}_k(\cdot)$ denote the conditional expectation given $\mathcal{F}_k$ for $k = 1, \cdots, n_x + n_y$. Define $D_{n,k} = (\mathrm{E}_k - \mathrm{E}_{k-1})Z_n$ and $\pi_{n,k}^2 = \mathrm{E}_{k-1}(D_{n,k}^2)$. It follows that $Z_n = \sum_{k=1}^{n} D_{n,k}$ as $\mathrm{E}_0(Z_n) = \mathrm{E}(Z_n) = 0$. To prove (B.24), by the martingale central limit theorem, it suffices to prove

$$\sum_{k=1}^{n} \pi_{n,k}^2/\mathrm{var}(Z_n) \xrightarrow{P} 1 \quad \text{and} \quad \sum_{k=1}^{n} \mathrm{E}(D_{n,k}^4)/\mathrm{var}^2(Z_n) \to 0. \qquad \text{(B.25)}$$

To prove (B.25), we derive the explicit forms of $D_{n,k}$ and $\pi_{n,k}^2$ in Section B.5.25. Similarly to Section B.1.1, the following Lemma B.3.3 and Lemma B.3.4 suggest that (B.25) holds.

Lemma B.3.3. *Under the conditions of Theorem 3.4.1,* $\mathrm{var}(\sum_{k=1}^{n_x+n_y} \pi_{n,k}^2) \to 0$.

Proof. *See Section B.5.26 on Page 413.* □

Lemma B.3.4. *Under the conditions of Theorem 3.4.1,* $\sum_{k=1}^{n_x+n_y} \mathrm{E}(D_{n,k}^4) \to 0$.

Proof. *See Section B.5.27 on Page 421.* □

In summary, Theorem 3.4.1 is proved.

B.3.2 Proof of Theorem 3.4.2

In this section, we first provide the conditions of Theorem 3.4.2 in Section B.3.2.1 and then prove Theorem 3.4.2 in Section B.3.2.2.

B.3.2.1 Conditions

Theorem 3.4.2 is established under the following Conditions B.3.1 and B.3.2, where Condition B.3.1 is the same as Condition 3.4.1* (1)–(3).

Condition B.3.1.

(1) $n, p \to \infty$, *and* $n_x/n \to \gamma \in (0,1)$.

(2) $\lim_{p\to\infty} \max_{1\le j\le p} \mathrm{E}(x_j - \mu_j)^8 < \infty$; $\lim_{p\to\infty} \min_{1\le j\le p} \mathrm{E}(x_j - \mu_j)^2 > 0$; $\lim_{p\to\infty} \max_{1\le j\le p} \mathrm{E}(y_j - \nu_j)^8 < \infty$; *and* $\lim_{p\to\infty} \min_{1\le j\le p} \mathrm{E}(y_j - \nu_j)^2 > 0$.

(3) For $t = 3, 4, 6, 8$, *there exist* $\kappa_{x,t}, \kappa_{y,t} \ge 1$ *such that* $\Pi^x_{j_1,\ldots,j_t} = \kappa_{x,t}\Pi^0_{j_1,\ldots,j_t}$ *and* $\Pi^y_{j_1,\ldots,j_t} = \kappa_{y,t}\Pi^0_{j_1,\ldots,j_t}$.

To provide Condition B.3.2, we first define some notation. The difference between $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ is defined as $D_{x,y} = \mathbf{\Sigma}_x - \mathbf{\Sigma}_y = (D_{j_1,j_2})_{p\times p}$. Let $\mathbb{J}_0 \subseteq \{1,\ldots,p\}$ be the largest set such that for any $j_1, j_2 \in \mathbb{J}_0$, $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2}$. Define $J_{0,D} = \{(j_1,j_2) : j_1 \text{ or } j_2 \notin \mathbb{J}_0\}$. Given $\mathbb{J}_0$ and $a, b \in \{a_1,\ldots,a_m\}$, we define $\mathbb{V}_{a,b,0,0} = \sum_{j_1,\ldots,j_8\in\mathbb{J}_0} (\sigma_{x,j_1,j_2}\sigma_{x,j_3,j_4})^a (\sigma_{x,j_5,j_6}\sigma_{x,j_7,j_8})^b$, which equals to $\sum_{j_1,\ldots,j_8\in\mathbb{J}_0} (\sigma_{y,j_1,j_2}\sigma_{y,j_3,j_4})^a \times (\sigma_{y,j_5,j_6}\sigma_{y,j_7,j_8})^b$ by the definition of $\mathbb{J}_0$. In addition, for any tuple $\mathcal{G} = (g_1, g_2, \ldots, g_{4(a+b)-1}, g_{4(a+b)}) \in \mathbb{G}_{a,b}$ specified in Condition 3.4.1*, we define $\mathbb{V}_{a,b,\mathcal{G},0} = \sum_{j_1,\ldots,j_8\in\mathbb{J}_0} \prod_{t=1}^{2(a+b)} \sigma_{j_{g_{2t-1}}, j_{g_{2t}}}$. Note that $\mathbb{V}_{a,b,0,0}$ and $\mathbb{V}_{a,b,\mathcal{G},0}$ are defined similarly to $\mathbb{V}_{a,b,0}$ and $\mathbb{V}_{a,b,\mathcal{G}}$ in Condition 3.4.1* by changing the range of j indexes from $\{1,\ldots,p\}$ to $\mathbb{J}_0$. Moreover, let $\mathcal{H} = \{(h_1,h_2),(h_3,h_4)\} \in \mathbb{H}$, where $\mathbb{H}$ includes $\{(1,2),(3,4)\}$, $\{(1,3),(2,4)\}$ and $\{(1,4),(2,3)\}$. For any $a \in \{a_1,\ldots,a_m\}$ and given $\mathcal{H} \in \mathbb{H}$, define

$$
\begin{aligned}
\mathbb{V}_{a,\mathcal{H},x,1} &= \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |\sigma_{x,j_{h_1},j_{h_2}}\sigma_{x,j_{h_3},j_{h_4}}|^a \\
\mathbb{V}_{a,\mathcal{H},x,2} &= \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |D_{j_{h_1},j_{h_2}}\sigma_{x,j_{h_3},j_{h_4}}|^a, \\
\mathbb{V}_{a,\mathcal{H},D,3} &= \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |D_{j_{h_1},j_{h_2}}D_{j_{h_3},j_{h_4}}|^a.
\end{aligned}
\tag{B.26}
$$

Similarly, we also define $\mathbb{V}_{a,\mathcal{H},y,1}$ and $\mathbb{V}_{a,\mathcal{H},y,2}$ by replacing σ_x's with σ_y's. We next present Condition B.3.2 of Theorem 3.4.2.

Condition B.3.2. *For any $a, b \in \{a_1, \ldots, a_m\}$, $\mathcal{G} \in \mathbb{G}_{a,b}$, and $\mathcal{H} \in \mathbb{H}$, we assume (A1) $\mathbb{V}_{a,b,\mathcal{G},0} = o(1)\mathbb{V}_{a,b,0,0}$; (A2) $\mathbb{V}_{a,\mathcal{H},D,3} = O(n^{-a})\mathbb{V}^{1/2}_{a,a,0,0}$; and (A3) $\mathbb{V}_{a,\mathcal{H},x,t} = o(1)\mathbb{V}^{1/2}_{a,a,0,0}$, for $t = 1, 2$.*

Equivalently we can also replace (A3) in Condition B.3.2 by (A3)* $\mathbb{V}_{a,\mathcal{H},y,t} = o(1)\mathbb{V}^{1/2}_{a,a,0,0}$, for $t = 1, 2$. This is because by $D_{j_1,j_2} = \sigma_{x,j_1,j_2} - \sigma_{y,j_1,j_2}$ and Hölder's inequality, we know *(A2)* and *(A3)* induce *(A3)**; and *(A2)* and *(A3)** also induce *(A3)*. Thus it is equivalent to assume *(A3)* or *(A3)** in Condition B.3.2.

We next discuss Condition B.3.2. Let $\mathbf{\Sigma}_C = \{\sigma_{x,j_1,j_2} : j_1, j_2 \in \mathbb{J}_0\} = \{\sigma_{y,j_1,j_2} : j_1, j_2 \in \mathbb{J}_0\}$, which is the common submatrix of $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ by the definition of $\mathbb{J}_0$. In Condition B.3.2, *(A1)* implies some weak dependence structure of $\mathbf{\Sigma}_C$ similar to Condition 3.4.1* (4). We consider an example where $\mathbf{\Sigma}_x$ has the banded structure with the bandwidth s and the entries being positive constants. Then *(A1)* holds if $s = o(p)$. Moreover, under the considered example, $\mathbb{V}^{1/2}_{a,a,0,0} = (\sum_{j_1,j_2 \in \mathbb{J}_0} \sigma^a_{x,j_1,j_2})^2 \geq C|\mathbb{J}_0|^4$ and $\mathbb{V}_{a,\mathcal{H},x,1} \leq C|J_{0,D}|^2 = C_2(p - |\mathbb{J}_0|)^4$. Then *(A3)* for $t = 1$ holds when $p - |\mathbb{J}_0| = o(p)$, which implies that the number of entries that are different in $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_y$ is $o(p^2)$. In addition, *(A2)* and *(A3)* for $t = 2$ are regularity conditions on the difference matrix $D_{x,y}$. For illustration, we consider an example where $D_{j_1,j_2} = \rho > 0$ for any $(j_1, j_2) \in J_{0,D}$, and $\mathbf{\Sigma}_x = I_p$. Then $\mathbb{V}^{1/2}_{a,a,0,0} = |\mathbb{J}_0|^2$, $\mathbb{V}_{a,\mathcal{H},x,2} \leq |J_{0,D}|\rho^a p$, and $\mathbb{V}_{a,\mathcal{H},D,3} \leq |J_{0,D}|^2 \rho^{2a}$. Under this example, *(A2)* and *(A3)* of $t = 2$ hold if $|J_{0,D}|\rho^a = O(n^{-a/2}p)$ and $|\mathbb{J}_0| \simeq p$, which are similar to the assumption in Theorem 3.2.5.

B.3.2.2 Proof of Theorem 3.4.2

In this section, we prove Theorem 3.4.2 under Conditions B.3.1 and B.3.2. Recall that we decompose $\mathcal{U}(a) = \tilde{\mathcal{U}}(a) + \tilde{\mathcal{U}}^*(a)$ in Section B.3.1. We further decompose

$\tilde{\mathcal{U}}(a) = T_{D,a,1} + T_{D,a,2}$, where

$$T_{D,a,1} = \sum_{j_1,j_2 \in \mathbb{J}_0} \frac{1}{P_a^{n_x} P_a^{n_y}} \sum_{\substack{1 \le i_1 \neq \ldots \neq i_a \le n_x; \\ 1 \le w_1 \neq \ldots \neq w_a \le n_y}} \prod_{t=1}^{a} (x_{i_t,j_1} x_{i_t,j_2} - y_{w_t,j_1} y_{w_t,j_2}),$$

$$T_{D,a,2} = \sum_{(j_1,j_2) \in J_{0,D}} \frac{1}{P_a^{n_x} P_a^{n_y}} \sum_{\substack{1 \le i_1 \neq \ldots \neq i_a \le n_x; \\ 1 \le w_1 \neq \ldots \neq w_a \le n_y}} \prod_{t=1}^{a} (x_{i_t,j_1} x_{i_t,j_2} - y_{w_t,j_1} y_{w_t,j_2}).$$

It follows that $\mathcal{U}(a) = T_{D,a,1} + T_{D,a,2} + \tilde{\mathcal{U}}^*(a)$. To prove Theorem 3.4.2, we derive the variances, covariances and asymptotic joint normality of the U-statistics. In particular, next Lemma B.3.5 derives the asymptotic form of $\mathrm{var}\{\mathcal{U}(a)\}$, and shows that $T_{D,a,1}$ is the leading component.

Lemma B.3.5. *Under the conditions of Theorem 3.4.2,*

$$\mathrm{var}\{\mathcal{U}(a)\} \simeq \sum_{1 \le j_1,j_2,j_3,j_4 \in \mathbb{J}_0} a! C_{\kappa,a} \sigma_{j_1,j_2}^a \sigma_{j_3,j_4}^a,$$

where $C_{\kappa,a} = \{(\kappa_x - 1)/n_x + (\kappa_y - 1)/n_y\}^a + 2(\kappa_x/n_x + \kappa_y/n_y)^a$. *In addition,* $\mathrm{var}(T_{D,a,2}) = o(1)\mathrm{var}(T_{D,a,1})$ *and* $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$. *It follows that* $\{T_{D,a,2} - \mathrm{E}(T_{D,a,2})\}/\sigma(a) \xrightarrow{P} 0$ *and* $[\tilde{\mathcal{U}}^*(a) - \mathrm{E}\{\tilde{\mathcal{U}}^*(a)\}]/\sigma(a) \xrightarrow{P} 0$.

Proof. *See Section B.5.28 on Page 424.* □

Lemma B.3.5 gives $\{T_{D,a,2} - \mathrm{E}(T_{D,a,2})\}/\sigma(a) \xrightarrow{P} 0$ and $[\tilde{\mathcal{U}}^*(a) - \mathrm{E}\{\tilde{\mathcal{U}}^*(a)\}]/\sigma(a) \xrightarrow{P} 0$. Thus by Slutsky's theorem, to prove Theorem 3.4.2, it suffices to prove

$$\Big[\frac{T_{D,a_1,1}}{\sqrt{\mathrm{var}(T_{D,a_1,1})}}, \ldots, \frac{T_{D,a_m,1}}{\sqrt{\mathrm{var}(T_{D,a_m,1})}}\Big] \xrightarrow{D} \mathcal{N}(\mathbf{0}, I_m). \tag{B.27}$$

Note that $T_{D,a,1}$ is a summation over j indexes in $\mathbb{J}_0$, and by the definition of $\mathbb{J}_0$, $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2}$ for any $j_1, j_2 \in \mathbb{J}_0$. Therefore the analysis under H_0 can be similarly applied to $T_{D,a,1}$. Given Condition B.3.1 and Condition B.3.2 *(A1)*, we can obtain (B.27) similarly as in Section B.3.1. In summary, Theorem 3.4.2 is proved.

B.3.3 Proof of Proposition 3.4.1

In this section, we prove Proposition 3.4.1. Under the considered example, as $p-|\mathbb{J}_0| = o(p)$, we have $\sum_{j_1,j_2,j_3,j_4\in\mathbb{J}_0} \sigma^a_{x,j_1,j_2}\sigma^a_{x,j_3,j_4} \simeq \{p\nu^{2a} + 2\sum_{t=1}^{s} h_t^a(p-t)\}^2$. Then by Lemma B.3.5, when $n_x = n_y = n/2$,

$$\mathrm{var}\{\mathcal{U}(a)\} \simeq (n/2)^{-a} a!(2\kappa_1^a + \kappa_2^a)\Big\{p\nu^{2a} + 2\sum_{t=1}^{s} h_t^a(p-t)\Big\}^2, \tag{B.28}$$

where $\kappa_1 = \kappa_x + \kappa_y$ and $\kappa_2 = \kappa_x + \kappa_y - 2$.

Recall that ρ_a is defined to be the value such that when $\rho = \rho_a$ under the alternative, $\mathrm{E}\{\mathcal{U}(a)\}/\sqrt{\mathrm{var}\{\mathcal{U}(a)\}} \simeq M$ for given M. By (B.28), ρ_a satisfies

$$|J_D|^2\rho_a^{2a} = M^2(n/2)^{-a}a!(2\kappa_1^a + \kappa_2^a)\Big\{p\nu^{2a} + 2\sum_{t=1}^{s} h_t^a(p-t)\Big\}^2.$$

We next obtain

$$\rho_a = \frac{(a!)^{\frac{1}{2a}}\sqrt{\kappa_1}\nu}{(n/2)^{1/2}}\Big(\frac{Mp}{|J_D|}\Big)^{1/a}\Big\{2 + \Big(\frac{\kappa_2}{\kappa_1}\Big)^a\Big\}^{\frac{1}{2a}}\Big\{1 + 2\sum_{t=1}^{s}\Big(\frac{h_t}{\nu^2}\Big)^a\Big(1-\frac{t}{p}\Big)\Big\}^{\frac{1}{a}}.$$

Let $\tilde{M} = Mp/|J_D|$, $\tilde{h}_t = h_t/\nu^2$, $\tilde{\nu} = \sqrt{\kappa_1}\nu$, and $\tilde{\kappa}_r = \kappa_2/\kappa_1$. It follows that

$$\rho_a = \tilde{\nu}(a!)^{\frac{1}{2a}}(n/2)^{-1/2}(\tilde{M})^{\frac{1}{a}}(2+\tilde{\kappa}_r^a)^{\frac{1}{2a}}\Big\{1 + 2\sum_{t=1}^{s}\tilde{h}_t^a\Big(1-\frac{t}{p}\Big)\Big\}^{\frac{1}{a}}.$$

Similarly to Section B.1.5, we study ρ_a as a function of integer a and show that if ρ_a starts to not decrease at some value, it will increase afterwards. Specifically, we show that when $\rho_{a+1}/\rho_a \geq 1$, $\rho_{a+2}/\rho_{a+1} > 1$. Note that

$$\begin{aligned}\frac{\rho_{a+1}}{\rho_a} &= \left[\frac{(a+1)!\tilde{M}^2(2+\tilde{\kappa}_r^{a+1})\Big\{1+2\sum_{t=1}^{s}\tilde{h}_t^{a+1}\Big(1-\frac{t}{p}\Big)\Big\}^2}{(a!)^{1+\frac{1}{a}}\tilde{M}^{2+\frac{2}{a}}(2+\tilde{\kappa}_r^a)^{1+\frac{1}{a}}\Big\{1+2\sum_{t=1}^{s}\tilde{h}_t^a\Big(1-\frac{t}{p}\Big)\Big\}^{2(1+\frac{1}{a})}}\right]^{\frac{1}{2(a+1)}}\\ &= \{\mathbb{D}(a)\tilde{M}^{-2}\}^{\frac{1}{2a(a+1)}},\end{aligned}$$

where $\mathbb{D}(a) = \mathbb{D}_1(a) \times \mathbb{D}_2(a) \times \mathbb{D}_3(a)$ with $\mathbb{D}_1(a) = (a+1)^a/a!$, $\mathbb{D}_2(a) = (2+\tilde{\kappa}_r^{a+1})^a/(2+\tilde{\kappa}_r^a)^{a+1}$ and

$$\mathbb{D}_3(a) = \Big\{1 + 2\sum_{t=1}^{s} \tilde{h}_t^{a+1}\Big(1 - \frac{t}{p}\Big)\Big\}^{2a} \Big/ \Big\{1 + 2\sum_{t=1}^{s} \tilde{h}_t^{a}\Big(1 - \frac{t}{p}\Big)\Big\}^{2(a+1)}.$$

It follows that $\rho_{a+1}/\rho_a > 1$ and $\rho_{a+1}/\rho_a = 1$ are equivalent to $\mathbb{D}(a) > \tilde{M}^2$ and $\mathbb{D}(a) = \tilde{M}^2$, respectively.

We next show that $\mathbb{D}(a)$ is a strictly increasing functions of a as $\mathbb{D}_1(a+1)/\mathbb{D}_1(a) > 1$, $\mathbb{D}_2(a+1)/\mathbb{D}_2(a) \geq 1$ and $\mathbb{D}_3(a+1)/\mathbb{D}_3(a) \geq 1$. Particularly,

$$\frac{\mathbb{D}_1(a+1)}{\mathbb{D}_1(a)} = \frac{(a+2)^{a+1}}{(a+1)!}\frac{a!}{(a+1)^a} = \Big(1 + \frac{1}{a+1}\Big)^{a+1} > 1;$$

$$\frac{\mathbb{D}_2(a+1)}{\mathbb{D}_2(a)} = \frac{(2+\tilde{\kappa}_r^{a+2})^{a+1}}{(2+\tilde{\kappa}_r^{a+1})^{a+2}} \times \frac{(2+\tilde{\kappa}_r^{a})^{a+1}}{(2+\tilde{\kappa}_r^{a+1})^{a}} = \Big\{\frac{(2+\tilde{\kappa}_r^{a+2})(2+\tilde{\kappa}_r^{a})}{(2+\tilde{\kappa}_r^{a+1})^2}\Big\}^{a+1} \geq 1,$$

where we use $2\tilde{\kappa}_r^{a+1} \leq \tilde{\kappa}_r^{a+2} + \tilde{\kappa}_r^{a}$ by the inequality of arithmetic and geometric means; and

$$\frac{\mathbb{D}_3(a+1)}{\mathbb{D}_3(a)} = \left[\frac{\Big\{1 + 2\sum_{t=1}^{s} \tilde{h}_t^{a+2}(1 - \frac{t}{p})\Big\}\Big\{1 + 2\sum_{t=1}^{s} \tilde{h}_t^{a}(1 - \frac{t}{p})\Big\}}{\Big\{1 + 2\sum_{t=1}^{s} \tilde{h}_t^{a+1}(1 - \frac{t}{p})\Big\}^2}\right]^{4(a+1)} \geq 1,$$

where we use $\sum_{t=1}^{s} \tilde{h}_t^{a+2}(1 - t/p) + \sum_{t=1}^{s} \tilde{h}_t^{a}(1 - t/p) \geq 2\sum_{t=1}^{s} \tilde{h}_t^{a+1}(1 - t/p)$ by the inequality of arithmetic and geometric means and $\{\sum_{t=1}^{s} \tilde{h}_t^{a+2}(1 - t/p)\}\{\sum_{t=1}^{s} \tilde{h}_t^{a}(1 - t/p)\} \geq \{\sum_{t=1}^{s} \tilde{h}_t^{a+1}(1 - t/p)\}^2$ by Hölder's inequality. In summary, $\mathbb{D}(a+1)/\mathbb{D}(a) > 1$, and thus $\mathbb{D}(a)$ is a strictly increasing function of a.

Given the monotonicity of $\mathbb{D}(a)$, we know that if $\mathbb{D}(a) \geq \tilde{M}^2$, $\mathbb{D}(a+1) > \tilde{M}^2$; equivalently this implies that if $\rho_{a+1} \geq \rho_a$, $\rho_{a+2} > \rho_{a+1}$. Suppose a_0 is the first integer such that $\mathbb{D}(a_0) \geq \tilde{M}^2$, i.e., for any integer $1 \leq a < a_0$, $\mathbb{D}(a) < \tilde{M}^2$. By the analysis above, we know ρ_a is decreasing when $a < a_0$, and ρ_a is strictly increasing when

$a > a_0$. Thus a_0 achieves the minimum of ρ_a. Since $\mathbb{D}(a)$ is strictly increasing in a, we know $a_0 < \infty$ given $\tilde{M}$, and a_0 increases as $\tilde{M}$ increases.

Moreover, as $s = o(p)$, there exists some constant C such that

$$\mathbb{D}(1) = \frac{2+\tilde{\kappa}_r^2}{(2+\tilde{\kappa}_r)^2} \times \frac{\{1+2\sum_{t=1}^{s} \tilde{h}_t^2(1-t/p)\}^2}{\{1+2\sum_{t=1}^{s} \tilde{h}_t(1-t/p)\}^4} \geq \mathbb{D}_0,$$

where

$$\mathbb{D}_0 = C \times \frac{2+\tilde{\kappa}_r^2}{(2+\tilde{\kappa}_r)^2} \times \frac{\{1+2\sum_{t=1}^{s} \tilde{h}_t^2\}^2}{\{1+2\sum_{t=1}^{s} \tilde{h}_t\}^4},$$

and we have $\mathbb{D}_0 = \Theta(1/s^2)$. Therefore, when $\mathbb{D}_0 \geq \tilde{M}^2$, i.e., $|J_D| \geq Mp/\sqrt{\mathbb{D}_0}$, we know $\mathbb{D}(1) \geq \tilde{M}^2$ and the minimum of $\mathbb{D}(a)$ is achieved at $a_0 = 1$. This indicates that the minimum of ρ_a is achieved at $a_0 = 1$.

B.4 Proofs of Theoretical Results in Section 3.5

B.4.1 Proof of Theorems 3.5.1 and 3.5.2

Theorem 3.5.1 is proved following the proof of Theorem 3.3.1 in Section B.2.1. Specifically, the arguments in Section B.2.1 can be applied to proving Theorem 3.5.1 by replacing $x_{i,j}$'s with $S_{i,j}$'s, and therefore the details are skipped.

The proof of Theorem 3.5.2 is similar to the proof of Theorem 3.3.5 in Section B.2.4. In particular, we decompose $\mathcal{U}(a) = T_{a,1} + T_{a,2}$, where we redefine

$$T_{a,1} = \sum_{j=1}^{k_0} \frac{1}{P_a^n} \sum_{1\leq i_1 \neq \cdots \neq i_a \leq n} \prod_{k=1}^{a} S_{i_k,j}, \quad T_{a,2} = \sum_{j=k_0+1}^{p} \frac{1}{P_a^n} \sum_{1\leq i_1 \neq \cdots \neq i_a \leq n} \prod_{k=1}^{a} S_{i_k,j}.$$

Note that $T_{a,2}$ is a summation over $j \in \{k_0+1,\ldots,p\}$ and $\mathrm{E}(S_j) = 0$ for $j \in \{k_0+1,\ldots,p\}$. Thus the conclusions similar to that in Theorem 3.5.1 hold for $T_{a,2}$.

Specifically, we have $\mathrm{var}(T_{a,2}) = \Theta\{(p-k_0)n^{-a}\}$ and

$$\left[T_{a_1,2}/\sqrt{\mathrm{var}(T_{a_1,2})}, \dots, T_{a_m,2}/\sqrt{\mathrm{var}(T_{a_m,2})}\right] \xrightarrow{D} \mathcal{N}(0, I_m). \tag{B.29}$$

When $\mathrm{var}(T_{a,1}) = o(1)\mathrm{var}(T_{a,2})$, which will be proved later, we have $\sigma^2(a) \simeq \mathrm{var}(T_{a,2})$ and $\{T_{a,1} - \mathrm{E}(T_{a,1})\}/\sigma(a) \xrightarrow{P} 0$. By the Slutsky's theorem and (B.29), Theorem 3.5.2 is proved.

To finish the proof of Theorem 3.5.2, it remains to prove $\mathrm{var}(T_{a,1}) = o(1)\mathrm{var}(T_{a,2})$. The analysis above gives that $\mathrm{var}(T_{a,2}) = \Theta\{(p-k_0)n^{-a}\}$. As $k_0 = o(p)$, to prove $\mathrm{var}(T_{a,1}) = o(1)\mathrm{var}(T_{a,2})$, it suffices to show $\mathrm{var}(T_{a,1}) = o(pn^{-a})$. Note that $\mathrm{var}(T_{a,1}) = \mathrm{E}(T_{a,1}^2) - \{\mathrm{E}(T_{a,1})\}^2$, $\mathrm{E}(T_{a,1}) = k_0\rho^a$, and

$$\mathrm{E}(T_{a,1}^2) = \frac{1}{(P_a^n)^2} \sum_{1\le j_1,j_2\le k_0} \sum_{\substack{1\le i_1\ne\cdots\ne i_a\le n;\\ 1\le \tilde{i}_1\ne\cdots\ne \tilde{i}_a\le n}} \mathrm{E}\Big\{\prod_{k=1}^{a}(S_{i_k,j_1}S_{\tilde{i}_k,j_2})\Big\}.$$

For $0 \le b \le a$, define an event $B_{S,b} = \{\{i_1,\dots,i_a\} \cap \{\tilde{i}_1,\dots,\tilde{i}_a\}$ is of size $b\}$ and correspondingly

$$G_{S,a,2,b} = (P_a^n)^{-2} \sum_{1\le j_1,j_2\le k_0} \sum_{\substack{1\le i_1\ne\cdots\ne i_a\le n;\\ 1\le \tilde{i}_1\ne\cdots\ne \tilde{i}_a\le n}} \mathrm{E}\Big\{\prod_{k=1}^{a}(S_{i_k,j_1}S_{\tilde{i}_k,j_2}) \times \mathbf{1}_{B_{S,b}}\Big\}.$$

Then $\mathrm{E}(T_{a,1}^2) = \sum_{b=0}^{a} G_{S,a,2,b}$. To prove $\mathrm{E}(T_{a,1}^2) - \{\mathrm{E}(T_{a,1})\}^2 = o(pn^{-a})$, we show $G_{S,a,2,0} - \{\mathrm{E}(T_{a,1})\}^2 = o(pn^{-a})$ and $\sum_{b=1}^{a} G_{S,a,2,b} = o(pn^{-a})$, respectively.

When $b=0$, $\{i_1,\dots,i_a\}\cap\{\tilde{i}_1,\dots,\tilde{i}_a\} = \emptyset$, and it follows that $G_{S,a,2,0} = (P_a^n)^{-2}k_0^2 \times P_{2a}^n\rho^{2a}$. By $\mathrm{E}(T_{a,1}) = k_0\rho^a$ and $k_0^2\rho^{2a} = O(pn^{-a})$, we have $|G_{S,a,2,0} - \{\mathrm{E}(T_{a,1})\}^2| = o(k_0^2\rho^{2a}) = o(pn^{-a})$. When $b \ge 1$,

$$G_{S,a,2,b} = C(P_a^n)^{-2} \sum_{1\le j_1,j_2\le k_0} P_{2a-b}^n(\sigma_{j_1,j_2} + \rho^2)^b \rho^{2(a-b)}.$$

The maximum order of $G_{S,a,2,b}$ is bounded by the following two quantities:

$$\sum_{1\le j_1,j_2\le k_0} \frac{P^n_{2a-b}}{(P^n_a)^2}\sigma^b_{j_1,j_2}\rho^{2(a-b)}, \tag{B.30}$$

$$\sum_{1\le j_1,j_2\le k_0} \frac{P^n_{2a-b}}{(P^n_a)^2}\rho^{2a}. \tag{B.31}$$

For (B.30), by $b \geq 1$, Condition 3.5.1 (3), and Lemma B.5.1, (B.30) $= O\{k_0 n^{-b}\rho^{2(a-b)}\}$. As $k_0 = o(p)$ and $\rho = O(k_0^{-1/a}p^{1/(2a)}n^{-1/2})$, we know (B.30) $= o(pn^{-a})$. For (B.31), when $b \geq 1$, (B.31) $= O(k_0^2 n^{-b}\rho^{2a}) = o(k_0^2\rho^{2a}) = o(pn^{-a})$. In summary, we have $|\mathrm{var}(T_{a,1})| \leq |\{\mathrm{E}(T_{a,1})\}^2 - G_{S,a,2,0}| + \sum_{b=1}^{a} |G_{S,a,2,b}| = o(pn^{-a})$. Therefore, Theorem 3.5.2 is proved.

B.5 Proofs of Technical Lemmas in Appendix B

B.5.1 Notation and Four Technical Lemmas

To facilitate the presentation of the proofs, we first introduce some notation and then provide four technical Lemmas B.5.1–B.5.4.

Notation We define some notation to simplify the representation of summations in the following proofs. For $a < n$, $\mathcal{P}(n,a)$ denotes the collection of a-tuples $\mathbf{i} = (i_1,\ldots,i_a)$ satisfying $1 \leq i_1 \neq \ldots \neq i_a \leq n$. Given $\mathbf{i} \in \mathcal{P}(n,a)$, we define $\{\mathbf{i}\}$ as the corresponding set containing the elements of $\mathbf{i}$ without order, that is, $\{\mathbf{i}\} = \{i_1,\ldots,i_a\}$. We apply usual set operations on the corresponding set of $\{\mathbf{i}\}$. For example, $|\{\mathbf{i}\}|$ denotes the size of the set $\{i_1,\ldots,i_a\}$, which is a in this case. In addition, for any two integers $a, b < n$, and two tuples $\mathbf{i} \in \mathcal{P}(n,a)$ and $\mathbf{m} \in \mathcal{P}(n,b)$, the operations $\{\mathbf{i}\} \cup \{\mathbf{m}\}$ and $\{\mathbf{i}\} \cap \{\mathbf{m}\}$ give the sets that equal to the union $\{i_1,\ldots,i_a\} \cup \{m_1,\ldots,m_b\}$ and intersection $\{i_1,\ldots,i_a\} \cap \{m_1,\ldots,m_b\}$ respectively. Moreover, we write $\{\mathbf{i}\} = \{\mathbf{m}\}$ and $\{\mathbf{i}\} \neq \{\mathbf{m}\}$ to indicate that the two sets $\{i_1,\ldots,i_a\}$

and $\{m_1, \ldots, m_b\}$ contain the same elements or not respectively.

In addition, let $\mathcal{C}(n, a)$ denote the collection of a-tuples $\mathbf{i} = (i_1, \ldots, i_a)$ satisfying $1 \le i_1, \ldots, i_a \le n$ without constraining the elements to be different. Similarly, we define $\{\mathbf{i}\}$ as the set containing the elements of $\mathbf{i}$ without order, and the set operations also apply similarly as above. Note that $|\{\mathbf{i}\}|$ may be smaller than a under this case.

We next list four technical lemmas which shall be used in the proofs later.

Lemma B.5.1. *(Guyon, 1995, Eq. (3.5)) Under the mixing assumption in Condition 3.2.2, suppose Z_1 and Z_2 are $\mathcal{Z}_1^t$-measurable and $\mathcal{Z}_{t+m}^{\infty}$-measurable random variables respectively. When $\mathrm{E}(|Z_1|^{2+\epsilon}) < \infty$ and $\mathrm{E}(|Z_2|^{2+\epsilon}) < \infty$, for some constants C and $\epsilon > 0$, $|\mathrm{cov}(Z_1, Z_2)| \le C\{\alpha(m)\}^{\frac{2}{2+\epsilon}}\{\mathrm{E}(|Z_1|^{2+\epsilon})\}^{\frac{1}{2+\epsilon}}\{\mathrm{E}(|Z_2|^{2+\epsilon})\}^{\frac{1}{2+\epsilon}}$.*

The lemma above can also be obtained from Lemma 2.4 in Kim (1994) by taking $p = q = 2 + \epsilon$.

Lemma B.5.2. *(Durrett, 2019, Lemma 3.4.3) When $|a_i| \le A$ and $|b_i| \le A$, then $|\prod_{i=1}^q a_i - \prod_{i=1}^q b_i| \le \sum_{i=1}^q |a_i - b_i| \, A^{q-1}$.*

Lemma B.5.3. *(Cai et al., 2013, Eq. (24)) for two series of numbers A_j and B_j for $j = 1, \ldots, p$. $|\max_{1\le j\le p} A_j^2 - \max_{1\le j\le p} B_j^2| \le 2\max_{1\le j\le p}|B_j|\max_{1\le j\le p}|A_j - B_j| + \max_{1\le j\le p}|A_j - B_j|^2$.*

Lemma B.5.4. *When $u, v \ge 0$ and $0 < \vartheta \le 1$, $(u+v)^{\vartheta} \le u^{\vartheta} + v^{\vartheta}$.*

Proof. When $u \ge 0$ and $0 < \vartheta \le 1$, $f(u) = u^{\vartheta}$ is concave function with $f(0) = 0$. By the subadditivity property of concave function, we have $f(u+v) \le f(u) + f(v)$. $\square$

B.5.2 Proof of Lemma B.1.1

To illustrate the main idea of the proof of Lemma B.1.1, we first consider a setting where $x_{i,j}$'s are all independent, and under this independence case we prove Lemma B.1.1 in Section B.5.2. Next in Section B.5.2, we prove Lemma B.1.1 under the dependence case with Condition 3.2.2. Last in Section B.5.2, we present the proof under Condition 3.2.2*

Proof illustration In this section, we present the proof of Lemma B.1.1 by only replacing Condition 3.2.2 with the assumption that $x_{i,j}$'s are independent. Recall $\tilde{\mathcal{U}}(a)$ defined in (3.5) and $\tilde{\mathcal{U}}^*(a) = \mathcal{U}(a) - \tilde{\mathcal{U}}(a)$. Then $\mathrm{var}\{\mathcal{U}(a)\} \leq \mathrm{var}\{\tilde{\mathcal{U}}(a)\} + 2\sqrt{\mathrm{var}\{\tilde{\mathcal{U}}(a)\}\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}} + \mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$. To prove Lemma B.1.1, we derive $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ and show $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$.

We derive $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ first. Under H_0, $\mathrm{E}(x_{i,j_1}x_{i,j_2}) = 0$ when $j_1 \neq j_2$. It follows that $\mathrm{E}\{\tilde{\mathcal{U}}(a)\} = 0$ and $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \mathrm{E}[\{\tilde{\mathcal{U}}(a)\}^2]$, and then

$$\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \frac{1}{(P_a^n)^2} \sum_{\substack{1\leq j_1\neq j_2\leq p \\ 1\leq j_3\neq j_4\leq p}} \sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a)} \mathrm{E}\Big\{\prod_{k=1}^{a}(x_{i_k,j_1}x_{i_k,j_2})(x_{\tilde{i}_k,j_3}x_{\tilde{i}_k,j_4})\Big\},$$

where following the notation defined in Section B.5.1, $\mathbf{i}$ and $\tilde{\mathbf{i}}$ represent some tuples $\mathbf{i} = (i_1, \dots, i_a)$ satisfying $1 \leq i_1 \neq \dots \neq i_a \leq n$; and $\tilde{\mathbf{i}} = (i_1, \dots, i_a)$ satisfying $1 \leq i_1 \neq \dots \neq i_a \leq n$. When the corresponding two sets $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$, for example, when index $i_1 \in \{\mathbf{i}\}$ but $i_1 \notin \{\tilde{\mathbf{i}}\}$,

$$\begin{aligned} &\mathrm{E}\Big\{\prod_{k=1}^{a}(x_{i_k,j_1}x_{i_k,j_2})(x_{\tilde{i}_k,j_3}x_{\tilde{i}_k,j_4})\Big\} \qquad \text{(B.32)} \\ &= \mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}) \times \mathrm{E}(\text{all the remaining terms}) = 0. \end{aligned}$$

Therefore, (B.32) $\neq 0$ only when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, i.e., $\{i_1, \dots, i_a\} = \{\tilde{i}_1, \dots, \tilde{i}_a\}$. In particular, when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$,

$$\mathrm{E}\Big\{\prod_{k=1}^{a}(x_{i_k,j_1}x_{i_k,j_2})(x_{\tilde{i}_k,j_3}x_{\tilde{i}_k,j_4})\Big\} = \{\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4})\}^a.$$

It follows that

$$\begin{aligned}\operatorname{var}\{\tilde{\mathcal{U}}(a)\} &= \frac{a!}{(P_a^n)^2}\sum_{\mathbf{i}\in\mathcal{P}(n,a)}\sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}}\{\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4})\}^a\\ &= \frac{a!}{P_a^n}\sum_{1\le j_1\ne j_2\le p;\,1\le j_3\ne j_4\le p}\{\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4})\}^a.\end{aligned}$$

When $x_{i,j}$'s are independent, as $j_1\ne j_2$ and $j_3\ne j_4$, $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4})\ne 0$ only when $\{j_1,j_2\}=\{j_3,j_4\}$, which gives $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4})=\mathrm{E}(x_{1,j_1}^2)\times\mathrm{E}(x_{1,j_2}^2)$. Therefore, $\operatorname{var}\{\tilde{\mathcal{U}}(a)\}=2a!(P_a^n)^{-1}\sum_{1\le j_1\ne j_2\le p}\mathrm{E}(x_{1,j_1}^2)\mathrm{E}(x_{1,j_2}^2)$. By Condition 3.2.1, we have $\operatorname{var}\{\tilde{\mathcal{U}}(a)\}=\Theta(p^2n^{-a})$.

We next show $\operatorname{var}\{\tilde{\mathcal{U}}^*(a)\}=o(1)\operatorname{var}\{\tilde{\mathcal{U}}(a)\}$. As $\mathrm{E}\{\tilde{\mathcal{U}}^*(a)\}=0$, $\operatorname{var}\{\tilde{\mathcal{U}}^*(a)\}=\mathrm{E}[\{\tilde{\mathcal{U}}^*(a)\}^2]$. Recall the definition of $\mathcal{U}^*(a)$, then we have

$$\operatorname{var}\{\tilde{\mathcal{U}}^*(a)\}=\sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}}\sum_{1\le c_1,c_2\le a}\sum_{\substack{\mathbf{i}\in\mathcal{P}(n,a+c_1)\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}}\frac{(-1)^{c_1+c_2}\binom{a}{c_1}\binom{a}{c_2}}{P_{a+c_1}^nP_{a+c_2}^n}Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4),$$

where we correspondingly define

$$\begin{aligned}Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)=\mathrm{E}\Bigg[&\prod_{k=1}^{a-c_1}x_{i_k,j_1}x_{i_k,j_2}\prod_{k=a-c_1+1}^{a}x_{i_k,j_1}\prod_{k=a+1}^{a+c_1}x_{i_k,j_2}\\ &\times\prod_{\tilde{k}=1}^{a-c_2}x_{\tilde{i}_{\tilde{k}},j_3}x_{\tilde{i}_{\tilde{k}},j_4}\prod_{\tilde{k}=a-c_2+1}^{a}x_{\tilde{i}_{\tilde{k}},j_3}\prod_{\tilde{k}=a+1}^{a+c_2}x_{\tilde{i}_{\tilde{k}},j_4}\Bigg].\end{aligned}$$

To evaluate $\operatorname{var}\{\tilde{\mathcal{U}}^*(a)\}$, we examine the value of $Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)$. We first note that if $Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\ne 0$, the following two claims hold:

Claim 1: $\{j_1,j_2\}=\{j_3,j_4\}$; *Claim 2:* $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$ and $c_1=c_2$.

To prove *Claim 1*, we show that if $\{j_1,j_2\}\ne\{j_3,j_4\}$, $Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)=0$. We consider $j_1\notin\{j_3,j_4\}$ as an example. When $j_1\notin\{j_3,j_4\}$, as $j_1\ne j_2$, we further know

$j_1 \notin \{j_2, j_3, j_4\}$ and we can write

$$Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}\Big(\prod_{k=1}^{a} x_{i_k, j_1}\Big) \times \mathrm{E}(\text{other terms with subscripts } j_2, j_3, j_4) = 0,$$

where we use $\mathrm{E}(\prod_{k=1}^{a} x_{i_k,j_1}) = \{\mathrm{E}(x_{1,j_1})\}^a = 0$ as $\mathrm{E}(x_{1,j_1}) = 0$. In addition, to prove *Claim 2*, we show that if $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$, $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = 0$. If $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$, similarly to (B.32), suppose an index $i \in \{\mathbf{i}\}$ but $i \notin \{\tilde{\mathbf{i}}\}$. Then we can write $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_1, j_2) = \mathrm{E}(x_{i,j_1}) \times \mathrm{E}(\text{other terms}) = 0$ or $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_1, j_2) = \mathrm{E}(x_{i,j_1} x_{i,j_2}) \times \mathrm{E}(\text{other terms}) = 0$. As $\{\mathbf{i}\}$ and $\{\tilde{\mathbf{i}}\}$ are of sizes $a + c_1$ and $a + c_2$ respectively, $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ induces $c_1 = c_2$.

Given *Claim 1* and *Claim 2*, we write $c_1 = c_2 = c$ and decompose $\{\mathbf{i}\}$ and $\{\tilde{\mathbf{i}}\}$ into three disjoint subsets respectively as follows:

$$\begin{aligned}
&\{\mathbf{i}\}_{(1)} = \{i_1, \ldots, i_{a-c}\},\ \{\mathbf{i}\}_{(2)} = \{i_{a-c+1}, \ldots, i_a\},\ \{\mathbf{i}\}_{(3)} = \{i_{a+1}, \ldots, i_{a+c}\}, \\
&\{\tilde{\mathbf{i}}\}_{(1)} = \{\tilde{i}_1, \ldots, \tilde{i}_{a-c}\},\ \{\tilde{\mathbf{i}}\}_{(2)} = \{\tilde{i}_{a-c+1}, \ldots, \tilde{i}_a\},\ \{\tilde{\mathbf{i}}\}_{(3)} = \{\tilde{i}_{a+1}, \ldots, \tilde{i}_{a+c}\},
\end{aligned}$$

which satisfies that $\{\mathbf{i}\} = \cup_{l=1}^{3} \{\mathbf{i}\}_{(l)}$ and $\{\tilde{\mathbf{i}}\} = \cup_{l=1}^{3} \{\tilde{\mathbf{i}}\}_{(l)}$. We next prove the following *Claim 3*: if $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \neq 0$, one of the following two cases hold:

1. $j_1 = j_3$, $j_2 = j_4$, $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$, $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(2)}$, $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(3)}$;

2. $j_1 = j_4$, $j_2 = j_3$, $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$, $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(3)}$, $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)}$.

To prove *Claim 3*, we note that *Claim 1* suggests that if $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \neq 0$, either $\{j_1 = j_3, j_2 = j_4\}$ or $\{j_1 = j_4, j_2 = j_3\}$ holds. We consider $j_1 = j_3$ and $j_2 = j_4$ as an example. Suppose that there exists an index $i \in \{\mathbf{i}\}_{(2)}$. Since $x_{i,j}$'s are independent with mean 0, if $i \in \{\tilde{\mathbf{i}}\}_{(1)}$, $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_1, j_2) = \mathrm{E}(x_{i,j_1}^2 x_{i,j_2}) \times \mathrm{E}(\text{other terms}) = 0$; or if $i \in \{\tilde{\mathbf{i}}\}_{(3)}$, $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_1, j_2) = \mathrm{E}(x_{i,j_1} x_{i,j_2}) \times \mathrm{E}(\text{other terms}) = 0$. Symmetrically, if $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_1, j_2) \neq 0$, we know $\{\mathbf{i}\}_{(l)} = \{\tilde{\mathbf{i}}\}_{(l)}$ for $l = 1, 2, 3$ under this case. The similar analysis also applies to the second case in *Claim 3*. Moreover, under the two cases in *Claim 3*, we have $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \{\mathrm{E}(x_{1,j_1}^2 x_{1,j_2}^2)\}^{a-c} \{\mathrm{E}(x_{1,j_1}^2)\}^c \{\mathrm{E}(x_{1,j_2}^2)\}^c$.

In summary,

$$
\begin{aligned}
\operatorname{var}\{\tilde{\mathcal{U}}^*(a)\} &= \sum_{1\le j_1\ne j_2\le p}\sum_{c=1}^{a}\sum_{\mathbf{i}\in\mathcal{P}(n,a+c)}\frac{2(a-c)!c!c!}{(P^n_{a+c})^2}\{\mathrm{E}(x^2_{1,j_1})\mathrm{E}(x^2_{1,j_2})\}^a \\
&\le C\sum_{1\le j_1\ne j_2\le p}\sum_{c=1}^{a} n^{-(a+c)}\{\mathrm{E}(x^2_{1,j_1})\mathrm{E}(x^2_{1,j_2})\}^a,
\end{aligned}
$$

which is of order $O(p^2n^{-(a+1)})$. Since we have obtained that $\operatorname{var}\{\tilde{\mathcal{U}}(a)\} = \Theta(p^2n^{-a})$, then $\operatorname{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\operatorname{var}\{\tilde{\mathcal{U}}(a)\}$ is proved.

Proof under Condition 3.2.2 Section B.5.2 considers the case where $x_{i,j}$'s are independent. In this section, we further prove Lemma B.1.1 under Condition 3.2.2. We first explain the proof idea intuitively. Under Condition 3.2.2, $x_{i,j}$'s may be no longer independent, but the dependence between x_{i,j_1} and x_{i,j_2} degenerates exponentially with their distance $|j_1 - j_2|$. We expect that when $|j_1 - j_2|$ is large enough, x_{i,j_1} and x_{i,j_2} are "asymptotically independent". Specifically, we will introduce a threshold K_0 to be defined in (B.40) below. Then we will show that the majority of (x_{i,j_1}, x_{i,j_2}) pairs satisfy $|j_1 - j_2| > K_0$, and when $|j_1 - j_2| > K_0$, x_{i,j_1} and x_{i,j_2} are weakly dependent with similar properties to those under the independence case.

We next present the detailed proof under Condition 3.2.2. Under H_0, similarly to Section B.5.2, we have $\mathrm{E}\{\mathcal{U}(a)\} = 0$ and $\operatorname{var}\{\mathcal{U}(a)\} = \mathrm{E}\{\mathcal{U}^2(a)\}$. Then

$$
\mathrm{E}\{\mathcal{U}^2(a)\} = \sum_{\substack{1\le j_1\ne j_2\le p;\\ 1\le j_3\ne j_4\le p}}\ \sum_{\substack{0\le c_1,c_2\le a;\\ \mathbf{i}\in\mathcal{P}(n,a+c_1);\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}} F(c_1,c_2,a)\times Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4), \tag{B.33}
$$

where we define $F(c_1, c_2, a) = (-1)^{c_1+c_2}\binom{a}{c_1}\binom{a}{c_2}(P^n_{a+c_1}P^n_{a+c_2})^{-1}$, and recall

$$\begin{aligned} &Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \qquad\qquad \text{(B.34)} \\ =&\mathrm{E}\Big\{\prod_{k=1}^{a-c_1} x_{i_k,j_1}x_{i_k,j_2}\prod_{k=a-c_1+1}^{a} x_{i_k,j_1}\prod_{k=a+1}^{a+c_1} x_{i_k,j_2}\prod_{\tilde{k}=1}^{a-c_2} x_{\tilde{i}_{\tilde{k}},j_3}x_{\tilde{i}_{\tilde{k}},j_4}\prod_{\tilde{k}=a-c_2+1}^{a} x_{\tilde{i}_{\tilde{k}},j_3}\prod_{\tilde{k}=a+1}^{a+c_2} x_{\tilde{i}_{\tilde{k}},j_4}\Big\}. \end{aligned}$$

Similarly to Section B.5.2, to evaluate $\mathrm{var}\{\mathcal{U}(a)\}$, we next examine the value of $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)$ under different cases.

When $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$, we show (B.34) $= 0$, that is, *Claim 2* in Section B.5.2 also holds here. To see this, we assume without loss of generality that an index $i \in \{\mathbf{i}\}$ and $i \notin \{\tilde{\mathbf{i}}\}$. Then (B.34) takes one of the two following forms:

$$\text{(B.34)} = \mathrm{E}(x_{i,j_1}) \times \mathrm{E}(\text{all the remaining terms}) \quad (j_1 = 1, \ldots, p),$$

$$\text{(B.34)} = \mathrm{E}(x_{i,j_1}x_{i,j_2}) \times \mathrm{E}(\text{all the remaining terms}) \quad (1 \le j_1 \neq j_2 \le p).$$

Since $\mathrm{E}(x_{i,j_1}) = 0$ and $\mathrm{E}(x_{i,j_1}x_{i,j_2}) = 0$ under H_0, we know (B.34) $= 0$ when $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$. It follows that

$$\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}}\ \sum_{\substack{0\le c_1,c_2\le a;\\ \mathbf{i}\in\mathcal{P}(n,a+c_1);\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}} F(c_1, c_2, a)Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{\{\mathbf{i}\}\neq\{\tilde{\mathbf{i}}\}\}} = 0, \qquad \text{(B.35)}$$

where $\mathbf{1}_{\{\cdot\}}$ represents an indicator function.

When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, we know $c_1 = c_2$ and we write $c_1 = c_2 = c$. If $c = 0$,

$$Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},c=0\}} = \{\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})\}^a.$$

Then we have

$$\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}}\sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},c=0\}} \tag{B.36}$$
$$= \frac{1}{(P_a^n)^2}\sum_{\mathbf{i}\in\mathcal{P}(n,a)} a!\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}}\{\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})\}^a$$
$$= a!(P_a^n)^{-1}\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}}\{\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})\}^a.$$

If $c\ge 1$, for given $\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)$, we decompose the sets $\{\mathbf{i}\}$ and $\{\tilde{\mathbf{i}}\}$ into three disjoint sets respectively, defined as:

$$\{\mathbf{i}\}_{(1)}=\{i_1,\ldots,i_{a-c}\},\ \{\mathbf{i}\}_{(2)}=\{i_{a-c+1},\ldots,i_a\},\ \{\mathbf{i}\}_{(3)}=\{i_{a+1},\ldots,i_{a+c}\},$$
$$\{\tilde{\mathbf{i}}\}_{(1)}=\{\tilde{i}_1,\ldots,\tilde{i}_{a-c}\},\ \{\tilde{\mathbf{i}}\}_{(2)}=\{\tilde{i}_{a-c+1},\ldots,\tilde{i}_a\},\ \{\tilde{\mathbf{i}}\}_{(3)}=\{\tilde{i}_{a+1},\ldots,\tilde{i}_{a+c}\},$$

which satisfy that $\{\mathbf{i}\}=\cup_{l=1}^3\{\mathbf{i}\}_{(l)}$ and $\{\tilde{\mathbf{i}}\}=\cup_{l=1}^3\{\tilde{\mathbf{i}}\}_{(l)}$. The definitions are similarly used in Section B.5.2. We next examine the value of (B.34) by further discussing different cases.

Case 1 We consider the cases where $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$, $1\le c\le a-1$ and $\{\mathbf{i}\}_{(1)}=\{\tilde{\mathbf{i}}\}_{(1)}$. Then we have $\{\mathbf{i}\}_{(2)}\cup\{\mathbf{i}\}_{(3)}=\{\tilde{\mathbf{i}}\}_{(2)}\cup\{\tilde{\mathbf{i}}\}_{(3)}$. Note that here $\{\mathbf{i}\}_{(1)}=\{\tilde{\mathbf{i}}\}_{(1)}$ is assumed, and $\{\mathbf{i}\}_{(2)},\{\mathbf{i}\}_{(3)},\{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\tilde{\mathbf{i}}\}_{(3)}$ are all nonempty as $c\ge 1$. Similarly to *Claim 3* in Section B.5.2, we next prove that if (B.34) $\neq 0$, one of the following two cases holds:

$$\{\mathbf{i}\}_{(3)}=\{\tilde{\mathbf{i}}\}_{(3)},\{\mathbf{i}\}_{(2)}=\{\tilde{\mathbf{i}}\}_{(2)},j_1=j_3,j_2=j_4; \tag{B.37}$$
$$\{\mathbf{i}\}_{(3)}=\{\tilde{\mathbf{i}}\}_{(2)},\{\mathbf{i}\}_{(2)}=\{\tilde{\mathbf{i}}\}_{(3)},j_1=j_4,j_2=j_3.$$

We prove (B.37) by contradiction.

If $\{\mathbf{i}\}_{(2)}\cap\{\tilde{\mathbf{i}}\}_{(2)}\neq\emptyset$ and $\{\mathbf{i}\}_{(2)}\cap\{\tilde{\mathbf{i}}\}_{(3)}\neq\emptyset$, it means that $\{\mathbf{i}\}_{(2)}$ intersects with both $\{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\tilde{\mathbf{i}}\}_{(3)}$. Suppose $i_1 \in \{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(2)}$ and $i_2 \in \{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(3)}$. It follows that

$$(\text{B.34}) = \mathrm{E}(x_{i_1,j_1}x_{i_1,j_3}) \times \mathrm{E}(x_{i_2,j_1}x_{i_2,j_4}) \times \mathrm{E}(\text{all the remaining terms}).$$

As $j_3 \neq j_4$, $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_3}) \times \mathrm{E}(x_{i_2,j_1}x_{i_2,j_4}) = 0$ under H_0. Therefore (B.34) $= 0$. Similarly if $\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(2)} \neq \emptyset$ and $\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(3)} \neq \emptyset$, we know (B.34) $= 0$. The analysis shows that when (B.34) $\neq 0$, $\{\mathbf{i}\}_{(2)}$ only intersects with one of $\{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\tilde{\mathbf{i}}\}_{(3)}$. Symmetrically, $\{\mathbf{i}\}_{(3)}$ only intersects with another one of $\{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\tilde{\mathbf{i}}\}_{(3)}$. Since $|\{\mathbf{i}\}_{(2)}| = |\{\mathbf{i}\}_{(3)}| = |\{\tilde{\mathbf{i}}\}_{(2)}| = |\{\tilde{\mathbf{i}}\}_{(3)}|$, it remains to consider two cases $\{\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(3)}\}$ or $\{\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(3)}$ and $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)}\}$. To obtain (B.37), we next examine the two cases respectively.

If $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(2)}$ and $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(3)}$, suppose $i_1 \in \{\mathbf{i}\}_{(2)}$ and $i_2 \in \{\mathbf{i}\}_{(3)}$. Then as $\{\mathbf{i}\}_{(2)} \cap \{\mathbf{i}\}_{(3)} = \emptyset$,

$$(\text{B.34}) = \mathrm{E}(x_{i_1,j_1}x_{i_1,j_3}) \times \mathrm{E}(x_{i_2,j_2}x_{i_2,j_4}) \times \mathrm{E}(\text{all the remaining terms}),$$

which is nonzero only when $j_1 = j_3$ and $j_2 = j_4$. Similarly, if $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(3)}$ and $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)}$, (B.34) $\neq 0$ only when $j_1 = j_4$ and $j_2 = j_3$. In summary, if (B.34) $\neq 0$, (B.37) is obtained, and

$$\begin{aligned}
&Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\{\mathbf{i}\}_{(1)}=\{\tilde{\mathbf{i}}\}_{(1)},1\leq c\leq a-1\}} \\
=&Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{\{\mathbf{i}\}_{(1)}=\{\tilde{\mathbf{i}}\}_{(1)},1\leq c\leq a-1\}} \times \left(\mathbf{1}_{\left\{\substack{\{\mathbf{i}\}_{(2)}=\{\tilde{\mathbf{i}}\}_{(2)},\, j_1=j_3, \\ \{\mathbf{i}\}_{(3)}=\{\tilde{\mathbf{i}}\}_{(3)},\, j_2=j_4}\right\}} + \mathbf{1}_{\left\{\substack{\{\mathbf{i}\}_{(2)}=\{\tilde{\mathbf{i}}\}_{(3)},\, j_1=j_4, \\ \{\mathbf{i}\}_{(3)}=\{\tilde{\mathbf{i}}\}_{(2)},\, j_2=j_3}\right\}} \right).
\end{aligned}$$

In addition, under the two cases in (B.37), we have $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \{\mathrm{E}(x_{i,j_1}^2 x_{i,j_2}^2)\}^{a-c} \times$

$\{\mathrm{E}(x_{i,j_1}^2)\mathrm{E}(x_{i,j_2}^2)\}^c$. Therefore,

$$\begin{aligned}
&\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}}\ \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\left\{\substack{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\\ \{\mathbf{i}\}_{(1)}=\{\tilde{\mathbf{i}}\}_{(1)},\\ 1\le c\le a-1}\right\}} \qquad \text{(B.38)}\\
=&\sum_{\substack{1\le c\le a-1;\\ \mathbf{i}\in\mathcal{P}(n,a+c);\\ 1\le j_1\neq j_2\le p}} \frac{\binom{a}{c}^2 2(a-c)!c!c!}{(P^n_{a+c})^2}\{\mathrm{E}(x_{i,j_1}^2x_{i,j_2}^2)\}^{a-c}\{\mathrm{E}(x_{i,j_1}^2)\mathrm{E}(x_{i,j_2}^2)\}^c.\\
=&\sum_{c=1}^{a-1} O(p^2n^{-(a+c)}),
\end{aligned}$$

where the last equation uses Condition 3.2.1.

Case 2 We consider the cases when $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$, $1\le c\le a-1$, $\{\mathbf{i}\}_{(1)}\neq\{\tilde{\mathbf{i}}\}_{(1)}$ and $\{\mathbf{i}\}_{(1)}\cap\{\tilde{\mathbf{i}}\}_{(1)}\neq\emptyset$. Suppose that there exists an index $i_1\in\{\mathbf{i}\}_{(1)}\cap\{\tilde{\mathbf{i}}\}_{(1)}$. Since $\{\mathbf{i}\}_{(1)}\neq\{\tilde{\mathbf{i}}\}_{(1)}$ and $|\{\mathbf{i}\}_{(1)}|=|\{\tilde{\mathbf{i}}\}_{(1)}|$, there exists another index $i_2\in\{\mathbf{i}\}_{(1)}$ and $i_2\notin\{\tilde{\mathbf{i}}\}_{(1)}$. As $\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}$, we know $i_2\in\{\tilde{\mathbf{i}}\}_{(2)}\cup\{\tilde{\mathbf{i}}\}_{(3)}$. Without loss of generality, we assume $i_2\in\{\tilde{\mathbf{i}}\}_{(2)}$, then

$$\text{(B.34)}=\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}x_{i_1,j_4})\mathrm{E}(x_{i_2,j_1}x_{i_2,j_2}x_{i_2,j_3})\mathrm{E}(\text{other terms}). \qquad \text{(B.39)}$$

As $j_1\neq j_2$ and $j_3\neq j_4$ in summation, it suffices to discuss four sub-cases $\{j_1=j_3$ and $j_2=j_4\}$, $\{j_1=j_4$ and $j_2=j_3\}$, $\{j_1\neq j_3$ and $j_1\neq j_4\}$ and $\{j_2\neq j_3$ and $j_2\neq j_4\}$ under Case 2.

Case 2.1 If $j_1=j_3$ and $j_2=j_4$, (B.39) gives

$$\text{(B.34)}=\mathrm{E}(x_{i_1,j_1}^2x_{i_1,j_2}^2)\times\mathrm{E}(x_{i_2,j_1}^2x_{i_2,j_2})\times\mathrm{E}(\text{all the remaining terms}).$$

When $x_{i,j}$'s are independent as in Section B.5.2, we know $\mathrm{E}(x_{i_2,j_1}^2x_{i_2,j_2})=\mathrm{E}(x_{i_2,j_1}^2)\times\mathrm{E}(x_{i_2,j_2})=0$ and thus (B.34) $=0$. Alternatively, under Condition 3.2.2, (B.34) may no longer be 0 due to the dependence of $x_{i,j}$'s. But as discussed at the beginning of Section B.5.2, we expect that x_{i,j_1} and x_{i,j_2} are "asymptotically independent" as $|j_1 - j_2|$ increases, and thus we expect that (B.34) is close to 0 when $|j_1 - j_2|$ is large. To quantitatively evaluate (B.34) based on $|j_1 - j_2|$, we introduce a threshold K_0 below, and discuss the value of (B.34) when $|j_1 - j_2| > K_0$ and $|j_1 - j_2| \le K_0$, respectively.

Specifically, given δ in Condition 3.2.2 and positive constants μ and ϵ, we define

$$K_0 = -(2+\epsilon)(4+\mu)(\log p)/(\epsilon \log \delta). \tag{B.40}$$

When $|j_1 - j_2| > K_0$, by Conditions 3.2.1 and 3.2.2, we have

$$|(\text{B.34})| \le C \times |\mathrm{E}(x_{i_2,j_1}^2 x_{i_2,j_2})| = C \times |\mathrm{cov}(x_{i_2,j_1}^2, x_{i_2,j_2})| \le C\delta^{\frac{K_0\epsilon}{2+\epsilon}} = O(1)p^{-(4+\mu)},$$

where $|\mathrm{cov}(x_{i_2,j_1}^2, x_{i_2,j_2})| \le C\delta^{\frac{K_0\epsilon}{2+\epsilon}}$ holds by the α-mixing inequality in Lemma B.5.1. When $|j_1 - j_2| \le K_0$, by the uniform boundedness of moments from Condition 3.2.1, we have (B.34) $= O(1)$. To summarize, we define an event $S_{nem} = \{\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}, 1 \le c \le a-1, \{\mathbf{i}\}_{(1)} \ne \{\tilde{\mathbf{i}}\}_{(1)}, \{\mathbf{i}\}_{(1)} \cap \{\tilde{\mathbf{i}}\}_{(1)} \ne \emptyset\}$. Then

$$\begin{aligned}
&Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{S_{nem}, j_1=j_3, j_2=j_4\}} \\
&= Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \left(\mathbf{1}_{\left\{ \substack{S_{nem}, j_1=j_3, j_2=j_4, \\ |j_1-j_2|>K_0} \right\}} + \mathbf{1}_{\left\{ \substack{S_{nem}, j_1=j_3, j_2=j_4, \\ |j_1-j_2|\le K_0} \right\}} \right).
\end{aligned}$$

The analysis above gives $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{S_{nem}, j_1=j_3, j_2=j_4, |j_1-j_2|>K_0\}} = O(1)p^{-(4+\mu)}$ and $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{S_{nem}, j_1=j_3, j_2=j_4, |j_1-j_2|\le K_0\}} = O(1)$, respectively. Moreover, the total number of (j_1, j_2) pairs satisfying $|j_1 - j_2| \le K_0$ and $|j_1 - j_2| > K_0$ are $O(p^2)$

and $O(pK_0)$, respectively. Therefore,

$$\left| \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{S_{nem},j_1=j_3,j_2=j_4\}} \right| \tag{B.41}$$

$$\le \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} \left|F(c,c,a)\right| \times \mathbf{1}_{\{S_{nem},j_1=j_3,j_2=j_4\}}$$

$$\times \left\{ O(p^{-(4+\mu)})\mathbf{1}_{\{|j_1-j_2|>K_0\}} + C\times \mathbf{1}_{\{|j_1-j_2|\le K_0\}} \right\}$$

$$= \sum_{c=1}^{a-1} n^{-(a+c)} \left\{ O(1)p^2p^{-(4+\mu)} + O(1)pK_0 \right\} = o(p^2n^{-a}).$$

Case 2.2 If $j_1 = j_4$ and $j_2 = j_3$, similarly to *Case 2.1*, we have

$$\left| \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{S_{nem},j_1=j_4,j_2=j_3\}} \right| \tag{B.42}$$

$$= o(p^2n^{-a}).$$

Case 2.3 We discuss the cases where $j_1 \neq j_3$ and $j_1 \neq j_4$. If $x_{i,j}$'s are independent as in Section B.5.2, we know $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}x_{i_1,j_4}) = \mathrm{E}(x_{i_1,j_1})\mathrm{E}(\text{other terms}) = 0$; thus by (B.39), (B.34) $= 0$ under this setting. Similarly to *Case 2.1*, under Condition 3.2.2, (B.34) may be no longer 0, and we will discuss the value of (B.34) using the threshold K_0 in (B.40).

To evaluate (B.34), by (B.39), we examine $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}x_{i_1,j_4})$. Let $(\tilde{j}_1,\tilde{j}_2,\tilde{j}_3,\tilde{j}_4)$ be the ordered version of (j_1,j_2,j_3,j_4) satisfying $\tilde{j}_1 \le \tilde{j}_2 \le \tilde{j}_3 \le \tilde{j}_4$. Then we have $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}x_{i_1,j_4}) = \mathrm{E}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4})$. Under the considered cases where $j_1 \neq j_3$ and $j_1 \neq j_4$, at least one of the two equations, $\mathrm{E}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}) = 0$ and $\mathrm{E}(x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4}) = 0$, holds. Then $\mathrm{E}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4}) = \mathrm{cov}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}\ ,\ x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4})$.

We thus can write

$$
\begin{aligned}
|\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}x_{i_1,j_4})| &= |\mathrm{E}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4})| \\
&= |\mathrm{cov}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}\ ,\ x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4})| \\
&= |\mathrm{cov}(x_{i_1,\tilde{j}_1}\ ,\ x_{i_1,\tilde{j}_2}x_{i_1,\tilde{j}_3}x_{i_1,\tilde{j}_4})| \\
&= |\mathrm{cov}(x_{i_1,\tilde{j}_1}x_{i_1,\tilde{j}_2}x_{i_1,\tilde{j}_3}\ ,\ x_{i_1,\tilde{j}_4})|.
\end{aligned} \tag{B.43}
$$

We next discuss the value of (B.43) based on the the maximum distance between the indexes in $(\tilde{j}_1, \tilde{j}_2, \tilde{j}_3, \tilde{j}_4)$, which is defined as

$$
\kappa_m = \max\{|\tilde{j}_2 - \tilde{j}_1|, |\tilde{j}_3 - \tilde{j}_2|, |\tilde{j}_4 - \tilde{j}_3|\}. \tag{B.44}
$$

We evaluate (B.43) when $\kappa_m > K_0$ and $\kappa_m \leq K_0$, respectively. First, if $\kappa_m > K_0$, by $\mathrm{E}(\mathbf{x}) = \mathbf{0}$, Conditions 3.2.1, 3.2.2, and Lemma B.5.1, we have (B.43) $\leq C\delta^{\frac{K_0\epsilon}{2+\epsilon}} = O(p^{-(4+\mu)})$. If $\kappa_m \leq K_0$, by Condition 3.2.1, (B.43) $= O(1)$. It follows that $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{S_{nem}, j_1\neq j_3, j_1\neq j_4, \kappa_m > K_0\}} = O(p^{-(4+\mu)})$, and $Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{S_{nem}, j_1\neq j_3, j_1\neq j_4, \kappa_m \leq K_0\}} = O(1)$, where the event S_{nem} is defined in *Case 2.1.* Note that the total number of (j_1, j_2, j_3, j_4) tuples satisfying $\kappa_m > K_0$ and $\kappa_m \leq K_0$ are $O(p^4)$ and $O(pK_0^3)$, respectively. Thus

$$
\begin{aligned}
&\Bigg| \sum_{\substack{1\leq j_1\neq j_2\leq p;\\ 1\leq j_3\neq j_4\leq p}} \sum_{\substack{0\leq c\leq a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{S_{nem}, j_1\neq j_3, j_1\neq j_4\}} \Bigg| \quad \text{(B.45)} \\
\leq\ & \sum_{\substack{1\leq j_1\neq j_2\leq p;\\ 1\leq j_3\neq j_4\leq p}} \sum_{\substack{0\leq c\leq a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} |F(c,c,a)| \times \mathbf{1}_{\{S_{nem}, j_1\neq j_3, j_1\neq j_4\}} \\
&\qquad\qquad \times \Big[O(p^{-(4+\mu)})\mathbf{1}_{\{\kappa_m > K_0\}} + C \times \mathbf{1}_{\{\kappa_m \leq K_0\}} \Big] \\
=\ & \sum_{c=1}^{a-1} n^{-(a+c)}\{p^2 O(p^{-(4+\mu)}) + O(1)pK_0^3\} = o(p^2 n^{-a}).
\end{aligned}
$$

Case 2.4 If $j_2 \neq j_3$ and $j_2 \neq j_4$, similarly to *Case 2.3*, we have

$$\Big|\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}}\ \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{S_{nem},j_2\neq j_3,j_2\neq j_4\}}\Big| = o(p^2n^{-a}). \tag{B.46}$$

By (B.41), (B.42), (B.45), (B.46), and the definition of S_{nem}, we obtain

$$\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}}\ \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4) \times\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},1\le c\le a-1,\{\mathbf{i}\}_{(1)}\neq\{\tilde{\mathbf{i}}\}_{(1)},\{\mathbf{i}\}_{(1)}\cap\{\tilde{\mathbf{i}}\}_{(1)}\neq\emptyset\}} = o(p^2n^{-a}). \tag{B.47}$$

Case 3 We consider $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, $1 \le c \le a-1$, and $\{\mathbf{i}\}_{(1)} \cap \{\tilde{\mathbf{i}}\}_{(1)} = \emptyset$. Here $\{\mathbf{i}\}_{(1)}$ and $\{\tilde{\mathbf{i}}\}_{(1)}$ are not empty as $c \le a-1$. Suppose there exist $i_1 \in \{\mathbf{i}\}_{(1)}$ and $i_2 \in \{\tilde{\mathbf{i}}\}_{(1)}$ with $i_1 \neq i_2$. Since $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{i}\}_{(1)} \cap \{\tilde{\mathbf{i}}\}_{(1)} = \emptyset$, we know $i_1 \in \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$ and $i_2 \in \{\mathbf{i}\}_{(2)} \cup \{\mathbf{i}\}_{(3)}$. Without loss of generality, we assume $i_1 \in \{\tilde{\mathbf{i}}\}_{(2)}$ and $i_2 \in \{\mathbf{i}\}_{(2)}$, then

$$\text{(B.34)} = \mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}) \times \mathrm{E}(x_{i_2,j_3}x_{i_2,j_4}x_{i_2,j_1}) \times \mathrm{E}(\text{other terms}).$$

To evaluate (B.34), we examine $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3})\mathrm{E}(x_{i_2,j_3}x_{i_2,j_4}x_{i_2,j_1})$. As $\mathrm{E}(\mathbf{x}) = \mathbf{0}$, we can write

$$\begin{aligned}\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3}) = \mathrm{cov}(x_{i_1,j_1}\,,\, x_{i_1,j_2}x_{i_1,j_3}) &= \mathrm{cov}(x_{i_1,j_2}\,,\, x_{i_1,j_1}x_{i_1,j_3})\\ &= \mathrm{cov}(x_{i_1,j_3}\,,\, x_{i_1,j_1}x_{i_1,j_2}),\end{aligned}$$

and similarly,

$$\mathrm{E}(x_{i_2,j_3}x_{i_2,j_4}x_{i_2,j_1}) = \mathrm{cov}(x_{i_2,j_3}\ ,\ x_{i_2,j_4}x_{i_2,j_1}) = \mathrm{cov}(x_{i_2,j_4}\ ,\ x_{i_2,j_3}x_{i_2,j_1})$$
$$= \mathrm{cov}(x_{i_2,j_1}\ ,\ x_{i_2,j_3}x_{i_2,j_4}).$$

Recall κ_m in (B.44) and K_0 in (B.40). If $\kappa_m > K_0$, by Conditions 3.2.1 and 3.2.2, and Lemma B.5.1, we have

$$\Big|\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3})\mathrm{E}(x_{i_2,j_3}x_{i_2,j_4}x_{i_2,j_1})\Big| \le C\delta^{\frac{K_0\epsilon}{2+\epsilon}} = O(1)p^{-(4+\mu)}. \qquad \text{(B.48)}$$

If $\kappa_m \le K_0$, by Condition 3.2.1, $\mathrm{E}(x_{i_1,j_1}x_{i_1,j_2}x_{i_1,j_3})\mathrm{E}(x_{i_2,j_3}x_{i_2,j_4}x_{i_2,j_1}) = O(1)$. Note that the total number of (j_1, j_2, j_3, j_4) tuples satisfying $\kappa_m > K_0$ and $\kappa_m \le K_0$ are $O(p^4)$ and $O(pK_0^3)$, respectively. Therefore,

$$\begin{aligned}
&\Big|\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}}\ \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\left\{\substack{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\};\\ 1\le c\le a-1;\\ \{\mathbf{i}\}_{(1)}\cap\{\tilde{\mathbf{i}}\}_{(1)}=\emptyset}\right\}}\Big| \qquad \text{(B.49)}\\
\le\ & \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}}\ \sum_{\substack{0\le c\le a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} \Big|F(c,c,a)\Big| \times \mathbf{1}_{\left\{\substack{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\};1\le c\le a-1;\\ \{\mathbf{i}\}_{(1)}\cap\{\tilde{\mathbf{i}}\}_{(1)}=\emptyset}\right\}}\\
&\times\Big[Cp^{-(4+\mu)}\mathbf{1}_{\{\kappa_m>K_0\}} + C\mathbf{1}_{\{\kappa_m\le K_0\}}\Big]\\
=\ & \sum_{c=1}^{a-1} n^{-(a+c)}\{O(1)p^4p^{-(4+\mu)} + O(1)pK_0^3\} = o(p^2n^{-a}).
\end{aligned}$$

Case 4 When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $c = a$, we know $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)} = \emptyset$ and $\{\mathbf{i}\}_{(2)} \cup \{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$. Then similarly *Case 1*, we have

$$\Big|\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}}\ \sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},c=a\}}\Big| = o(p^2n^{-a}). \qquad \text{(B.50)}$$

In summary, by (B.33), (B.35)–(B.38), (B.47), (B.49), and (B.50),

$$\mathrm{var}\{\mathcal{U}(a)\} = \frac{a!}{P_a^n} \sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}} \{\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})\}^a + o(p^2 n^{-a}). \tag{B.51}$$

Note that we assume $\mathrm{E}(\mathbf{x}) = \mathbf{0}$. For the general case with $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$, by Proposition 3.2.1, it is equivalent to replace $x_{i,j}$ by $x_{i,j} - \mu_j$ in (B.51).

We next show that $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ = (B.36) and $\mathrm{var}[\tilde{\mathcal{U}}^*(a)] = o(p^2n^{-a})$. First note that $\mathrm{E}\{\tilde{\mathcal{U}}(a)\} = \mathrm{E}\{\tilde{\mathcal{U}}^*(a)\} = 0$ under H_0 as $\mathrm{E}(\mathbf{x}) = \mathbf{0}$. Then it suffices to show $\mathrm{E}\{\{\tilde{\mathcal{U}}(a)\}^2\}$ = (B.36) and $\mathrm{E}\{\{\tilde{\mathcal{U}}^*(a)\}^2\} = o(p^2n^{-a})$. By the definition of $\tilde{\mathcal{U}}(a)$ in (3.5), we know

$$\mathrm{E}\{\tilde{\mathcal{U}}^2(a)\} = \sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}} \sum_{\substack{0\le c_1,c_2\le a;\\ \mathbf{i}\in\mathcal{P}(n,a+c_1);\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}} F(c_1,c_2,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\times \mathbf{1}_{\{c_1=c_2=0\}}. \tag{B.52}$$

Therefore, $\mathrm{E}\{\tilde{\mathcal{U}}^2(a)\}$ = (B.36) from previous discussion. Moreover, as $\tilde{\mathcal{U}}^*(a) = \mathcal{U}(a) - \tilde{\mathcal{U}}(a)$, we know

$$\begin{aligned}\tilde{\mathcal{U}}^*(a) = &\sum_{c=0}^{a} \mathbf{1}_{\{c\ge1\}} \sum_{1\le j_1\ne j_2\le p} (-1)^c \binom{a}{c} \frac{1}{P_{a+c}^n} \sum_{\mathbf{i}\in\mathcal{P}(n,a+c)} \\ &\times \prod_{k=1}^{a-c}(x_{i_k,j_1}x_{i_k,j_2}) \prod_{k=a-c+1}^{a} x_{i_k,j_1} \prod_{k=a+1}^{a+c} x_{i_k,j_2}.\end{aligned} \tag{B.53}$$

It follows that

$$\begin{aligned}&\mathrm{E}[\{\tilde{\mathcal{U}}^*(a)\}^2] \\ = &\sum_{\substack{1\le j_1\ne j_2\le p\\ 1\le j_3\ne j_4\le p}} \sum_{\substack{0\le c_1,c_2\le a;\\ \mathbf{i}\in\mathcal{P}(n,a+c_1);\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}} F(c_1,c_2,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\times \mathbf{1}_{\{c_1\ge1,c_2\ge1\}}.\end{aligned} \tag{B.54}$$

Also by previous discussion, we know $\mathrm{E}[\{\tilde{\mathcal{U}}^*(a)\}^2] = o(p^2n^{-a})$.

To finish the proof of Lemma B.1.1, it remains to show $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \text{(B.36)} = \Theta(p^2 n^{-a})$, and it suffices to prove

$$\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \{\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})\}^a = \Theta(p^2). \tag{B.55}$$

To prove (B.55), we examine $\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})$. Similarly to *Case 2* above, as $j_1 \neq j_2$ and $j_3 \neq j_4$ in summation, it suffices to discuss four cases $\{j_1 = j_3 \text{ and } j_2 = j_4\}$, $\{j_1 = j_4 \text{ and } j_2 = j_3\}$, $\{j_1 \neq j_3 \text{ and } j_1 \neq j_4\}$, and $\{j_2 \neq j_3 \text{ and } j_2 \neq j_4\}$.

If $j_1 = j_3$, $j_2 = j_4$, and $|j_1 - j_2| > K_0$, then by Conditions 3.2.1, 3.2.2, and Lemma B.5.1, we have

$$\begin{aligned}|\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})| =&\mathrm{E}(x_{i,j_1}^2x_{i,j_2}^2) = \mathrm{cov}(x_{i,j_1}^2, x_{i,j_2}^2) + \mathrm{E}(x_{i,j_1}^2)\mathrm{E}(x_{i,j_2}^2)\\ \geq&\Theta(1) - |\mathrm{cov}(x_{i,j_1}^2, x_{i,j_2}^2)| \geq \Theta(1) - C\delta^{\frac{K_0\epsilon}{2+\epsilon}} = \Theta(1).\end{aligned}$$

If $j_1 = j_3$, $j_2 = j_4$, and $|j_1 - j_2| \leq K_0$, by Condition 3.2.1, $\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4}) = O(1)$. Note that (j_1, j_2) pairs satisfying $|j_1 - j_2| > K_0$ and $|j_1 - j_2| \leq K_0$ are $O(p^2)$ and $O(pK_0)$, respectively. Thus,

$$\begin{aligned}&\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} [\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})]^a \mathbf{1}_{\{j_1=j_3,j_2=j_4\}} \qquad\qquad \text{(B.56)}\\ =&\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \Big[\mathrm{E}\Big(\prod_{t=1}^4 x_{i,j_t}\Big)\Big]^a \mathbf{1}_{\left\{\substack{j_1=j_3,\\ j_2=j_4}\right\}}[\mathbf{1}_{\{|j_1-j_2|>K_0\}} + \mathbf{1}_{\{|j_1-j_2|\leq K_0\}}]\\ =&\,\Theta(p^2) + O(pK_0) = \Theta(p^2).\end{aligned}$$

If $j_1 = j_4$ and $j_2 = j_3$, similarly to (B.56), we have

$$\sum_{\substack{1\le j_1\neq j_2\le p\\ 1\le j_3\neq j_4\le p}} [\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})]^a \mathbf{1}_{\{j_1=j_4,j_2=j_3\}} = \Theta(p^2). \tag{B.57}$$

If $j_1 \neq j_3$ and $j_1 \neq j_4$, we know (B.43) holds. Recall K_0 in (B.40) and κ_m in (B.44). Similarly to the analysis of (B.45), we have

$$\begin{aligned}
&\sum_{\substack{1\le j_1\neq j_2\le p\\ 1\le j_3\neq j_4\le p}} [\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})]^a \mathbf{1}_{\{j_1\neq j_3, j_1\neq j_4\}} \\
&= \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \Big[\mathrm{E}\Big(\prod_{t=1}^{4} x_{i,j_t}\Big)\Big]^a \mathbf{1}_{\{j_1\neq j_3, j_1\neq j_4\}} \Big[\mathbf{1}_{\{\kappa_m > K_0\}} + \mathbf{1}_{\{\kappa_m \le K_0\}}\Big] \\
&= o(p^2).
\end{aligned} \tag{B.58}$$

If $j_2 \neq j_3$ and $j_2 \neq j_4$, similarly to (B.58), we have

$$\sum_{\substack{1\le j_1\neq j_2\le p\\ 1\le j_3\neq j_4\le p}} [\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})]^a \mathbf{1}_{\{j_2\neq j_3, j_2\neq j_4\}} = o(p^2). \tag{B.59}$$

In summary, combining (B.56)–(B.59), we have

$$\sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \Big[\mathrm{E}\Big(\prod_{t=1}^{4} x_{i,j_t}\Big)\Big]^a \simeq 2 \sum_{1\le j_1\neq j_2\le p} \{\mathrm{E}(x_{i,j_1}^2 x_{i,j_2}^2)\}^a. \tag{B.60}$$

Combining (B.51), (B.52) and (B.60), Lemma B.1.1 is proved.

Proof under Condition 3.2.2* In this section, we prove Lemma B.1.1 by substituting Condition 3.2.2 with Condition 3.2.2*. Following the notation in Section B.5.2, we have

$$\mathrm{var}\{\mathcal{U}(a)\} = \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{0\le c_1,c_2\le a;\\ \mathbf{i}\in\mathcal{P}(n,a+c_1);\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,a+c_2)}} F(c_1,c_2,a)\times Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4).$$

When $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$, under H_0, we know (B.34) $= 0$ and (B.35) holds similarly. As $\{\mathbf{i}\}$ and $\{\tilde{\mathbf{i}}\}$ are of sizes $a + c_1$ and $a + c_2$ respectively, in the following we consider

$\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, which induces $c_1 = c_2$ and we write $c_1 = c_2 = c$.

When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $c = 0$, we know (B.36) also holds similarly, and $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} =$ (B.36) by (B.52). By Condition 3.2.2*,

$$\begin{aligned} &\mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4}) \\ &= \kappa_1\Big\{\mathrm{E}(x_{i,j_1}x_{i,j_2})\mathrm{E}(x_{i,j_3}x_{i,j_4}) + \mathrm{E}(x_{i,j_1}x_{i,j_3})\mathrm{E}(x_{i,j_2}x_{i,j_4}) \\ &\quad + \mathrm{E}(x_{i,j_1}x_{i,j_4})\mathrm{E}(x_{i,j_2}x_{i,j_3})\Big\}. \end{aligned} \tag{B.61}$$

Since $j_1 \neq j_2$ and $j_3 \neq j_4$, we know under H_0, (B.61) $\neq 0$ only when $\{j_1 = j_3, j_2 = j_4\}$ or $\{j_1 = j_4, j_2 = j_3\}$; and then (B.61) $= \kappa_1\mathrm{E}(x_{i,j_1}^2)\mathrm{E}(x_{i,j_2}^2)$. Thus

$$\text{(B.36)} = 2a!(P_a^n)^{-1}\sum_{1\le j_1\neq j_2\le p}\{\kappa_1\mathrm{E}(x_{i,j_1}^2)\mathrm{E}(x_{i,j_2}^2)\}^a = \Theta(p^2n^{-a}),$$

where the second equation follows from Condition 3.2.1.

When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $c \ge 1$, $|\{\mathbf{i}\}_{(2)}| = |\{\mathbf{i}\}_{(3)}| = |\{\tilde{\mathbf{i}}\}_{(2)}| = |\{\tilde{\mathbf{i}}\}_{(3)}| > 0$. Without loss of generality, we first consider an index $i \in \{\mathbf{i}\}_{(2)}$, and discuss four cases.

Case 1.1 If $i \notin \{\tilde{\mathbf{i}}\}$, since $\mathrm{E}(\mathbf{x}) = \mathbf{0}$, we know

$$\text{(B.34)} = \mathrm{E}(x_{i,j_1}) \times \mathrm{E}(\text{all the remaining terms}) = 0.$$

Case 1.2 If $i \in \{\tilde{\mathbf{i}}\}_{(2)}$,

$$\text{(B.34)} = \mathrm{E}(x_{i,j_1}x_{i,j_3}) \times \mathrm{E}(\text{all the remaining terms}),$$

which is nonzero when $j_1 = j_3$.

Case 1.3 If $i \in \{\tilde{\mathbf{i}}\}_{(3)}$,

$$\text{(B.34)} = \mathrm{E}(x_{i,j_1}x_{i,j_4}) \times \mathrm{E}(\text{all the remaining terms}) = 0,$$

which is nonzero when $j_1 = j_4$.

Case 1.4 If $i \in \{\tilde{\mathbf{i}}\}_{(1)}$, this suggests $\{\mathbf{i}\}_{(1)} \neq \emptyset$ and thus $c \leq a-1$. By Condition 3.2.2*,

$$\text{(B.34)} = \mathrm{E}(x_{i,j_1}x_{i,j_3}x_{i,j_4}) \times \mathrm{E}[\text{all the remaining terms}] = 0. \qquad \text{(B.62)}$$

When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $c \leq a-1$, we have $\{\mathbf{i}\}_{(1)} \neq \emptyset$. We assume without loss of generality that an index $i \in \{\mathbf{i}\}_{(1)}$, and then discuss two cases.

Case 2.1 If $i \in \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$, symmetrically, (B.34) takes a form similarly to that in (B.62), which is 0 under H_0 by Condition 3.2.2*.

Case 2.2 If $i \notin \{\tilde{\mathbf{i}}\}$, by $j_1 \neq j_2$, we know under H_0,

$$\text{(B.34)} = \mathrm{E}(x_{i,j_1}x_{i,j_2}) \times \mathrm{E}(\text{all the remaining terms}) = 0.$$

In summary, (B.34) $\neq 0$ only when one of the following two cases holds:

1. $j_1 = j_3$, $j_2 = j_4$, $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$, $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(2)}$, $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(3)}$;

2. $j_1 = j_4$, $j_2 = j_3$, $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$, $\{\mathbf{i}\}_{(2)} = \{\tilde{\mathbf{i}}\}_{(3)}$, $\{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)}$.

Under these two cases, (B.34) $= \{\kappa_1 \mathrm{E}(x_{i,j_1}^2 x_{i,j_2}^2)\}^{a-c}\{\mathrm{E}(x_{i,j_1}^2)\}^c\{\mathrm{E}(x_{i,j_2}^2)\}^c$. It follows that when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $c \geq 1$,

$$\begin{aligned}
&\sum_{\substack{1\leq j_1\neq j_2\leq p;\\ 1\leq j_3\neq j_4\leq p}} \sum_{\substack{0\leq c\leq a;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)}} F(c,c,a)Q(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},c\geq 1\}} \qquad \text{(B.63)}\\
=& \sum_{\substack{1\leq c\leq a;\\ 1\leq j_1\neq j_2\leq p}} \binom{a}{c}^2 \frac{2}{P^n_{a+c}}\{\kappa_1\mathrm{E}(x_{i,j_1}^2 x_{i,j_2}^2)\}^{a-c}\{\mathrm{E}(x_{i,j_1}^2)\}^c\{\mathrm{E}(x_{i,j_2}^2)\}^c\\
=& \sum_{c=1}^{a} O(p^2 n^{-(a+c)}) = o(pn^{-a}),
\end{aligned}$$

where the last two equations use Condition 3.2.1. Similarly to Section B.5.2, by

(B.35) and (B.54), we know $\text{var}\{\tilde{\mathcal{U}}^*(a)\} = \text{(B.63)} = o(pn^{-a}) = o(1)\text{var}\{\tilde{\mathcal{U}}(a)\}$.

Remark B.1. *κ_1 is assumed to be a constant in Condition 3.2.2*. But the similar arguments apply in the proof if κ_1 changes with n, p but converges to a constant.*

B.5.3 Proof of Lemma B.1.2

Note that for two integers $a \neq b$, $\text{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} = \text{E}[\mathcal{U}(a)\mathcal{U}(b)/\{\sigma(a)\sigma(b)\}]$, and by Lemma B.1.1, $\text{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\text{var}\{\tilde{\mathcal{U}}(a)\}$. Recall $\sigma^2(a) = \text{var}\{\mathcal{U}(a)\}$ from definition. Then by Cauchy-Schwarz inequality, we have

$$\text{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} = \text{E}\{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)\}/\{\sigma(a)\sigma(b)\} + o(1).$$

In addition,

$$\text{E}\{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)\} = \sum_{\substack{1\le j_1\neq j_2\le p,\\ 1\le j_3\neq j_4\le p}} \sum_{\substack{\mathbf{i}\in\mathcal{P}(n,a),\\ \tilde{\mathbf{i}}\in\mathcal{P}(n,b)}} \text{E}\Big\{\prod_{k=1}^{a}(x_{i_k,j_1}x_{i_k,j_2})\prod_{\tilde{k}=1}^{b}(x_{\tilde{i}_{\tilde{k}},j_3}x_{\tilde{i}_{\tilde{k}},j_4})\Big\}.$$

Since $a \neq b$, we know the two sets $\{i_1, \ldots, i_a\}$ and $\{\tilde{i}_1, \ldots, \tilde{i}_b\}$ can not be the same. Following similar analysis to that of (B.32), as $\text{E}(x_{i,j_1}x_{i,j_2}) = 0$ under H_0, we have $\text{E}\{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)\} = 0$, and thus $\text{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} = o(1)$.

In particular, we note that given Lemma B.1.1, the argument does not depend on whether Condition 3.2.2 or 3.2.2* is specified.

B.5.4 Proof of Lemma B.1.3

We first show for $1 \le k_1 \neq k_2 \le n$, $\text{E}(D_{n,k_1}D_{n,k_2}) = 0$. Without loss of generality, we consider $k_1 < k_2$. Then $\text{E}_{k_1}Z_n \in \mathcal{F}_{k_2}$, and

$$\begin{aligned}&\text{E}(D_{n,k_1}D_{n,k_2})\\ =&\text{E}[(\text{E}_{k_1}Z_n)Z_n] - \text{E}[(\text{E}_{k_1-1}Z_n)Z_n] - \text{E}[(\text{E}_{k_1}Z_n)Z_n] + \text{E}[(\text{E}_{k_1-1}Z_n)Z_n] = 0.\end{aligned}$$

It follows that

$$\mathrm{E}\left(\sum_{k=1}^{n} \pi_{n,k}^2\right) = \sum_{k=1}^{n} \mathrm{E}\left(D_{n,k}^2\right) = \mathrm{E}\left(\sum_{k=1}^{n} D_{n,k}\right)^2 = \mathrm{var}(Z_n),$$

where the last equation uses the fact that $\mathrm{E}(D_{n,k}) = 0$ and $Z_n = \sum_{k=1}^{n} D_{n,k}$ from construction.

B.5.5 Proof of Lemma B.1.4

For given finite integer a, we derive the expression of $(\mathrm{E}_k - \mathrm{E}_{k-1})[\tilde{\mathcal{U}}(a)/\sigma(a)]$. The form of A_{n,k,a_r} for a general finite integer a_r in Lemma B.1.4 follows similarly.

By the definition in (3.5), we know

$$(\mathrm{E}_k - \mathrm{E}_{k-1})\tilde{\mathcal{U}}(a) = (P_a^n)^{-1} \sum_{\substack{1 \le j_1 \ne j_2 \le p; \\ \mathbf{i} \in \mathcal{P}(n,a)}} (\mathrm{E}_k - \mathrm{E}_{k-1})\Big[\prod_{t=1}^{a} x_{i_t,j_1} x_{i_t,j_2}\Big]. \tag{B.64}$$

To derive (B.64), we next examine the value of

$$(\mathrm{E}_k - \mathrm{E}_{k-1})\Big[\prod_{t=1}^{a} x_{i_t,j_1} x_{i_t,j_2}\Big]. \tag{B.65}$$

We claim (B.65) $\ne 0$ only when $k \in \{i_1, \ldots, i_a\}$. If $k \notin \{i_1, \ldots, i_a\}$, we assume without loss of generality that $i_1, \ldots, i_m < k$ and $i_{m+1}, \ldots, i_a > k$. Then

$$\begin{aligned}
&(\mathrm{E}_k - \mathrm{E}_{k-1})\Big[\prod_{t=1}^{a} x_{i_t,j_1} x_{i_t,j_2}\Big] \\
=& \Big(\prod_{t=1}^{m} x_{i_t,j_1} x_{i_t,j_2}\Big)\Big[\mathrm{E}_k\Big(\prod_{t=m+1}^{a} x_{i_t,j_1} x_{i_t,j_2}\Big) - \mathrm{E}_{k-1}\Big(\prod_{t=m+1}^{a} x_{i_t,j_1} x_{i_t,j_2}\Big)\Big] = 0.
\end{aligned}$$

Thus if (B.65) $\ne 0$, we know $k \in \{i_1, \ldots, i_a\}$. In addition, we next show (B.65) $\ne 0$ only when $i_1, \ldots, i_a \le k$. Suppose that if there exist some indexes in $\{i_1, \ldots, i_a\}$ that are greater than k, we assume without loss of generality that $i_m = k$, $i_1, \ldots, i_{m-1} < k$, and $i_{m+1}, \ldots, i_a > k$. Then

$$
\begin{aligned}
\mathrm{E}_k\Big(\prod_{t=1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) &= \Big(\prod_{t=1}^{m} x_{i_t,j_1}x_{i_t,j_2}\Big)\mathrm{E}_k\Big(\prod_{t=m+1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) \\
&= \Big(\prod_{t=1}^{m} x_{i_t,j_1}x_{i_t,j_2}\Big)\mathrm{E}\Big(\prod_{t=m+1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{E}_{k-1}\Big(\prod_{t=1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) &= \Big(\prod_{t=1}^{m-1} x_{i_t,j_1}x_{i_t,j_2}\Big)\mathrm{E}_{k-1}\Big(x_{k,j_1}x_{k,j_2}\prod_{t=m+1}^{a} x_{i_t,j_1}x_{i_t,j_2}\Big) \\
&= \Big(\prod_{t=1}^{m-1} x_{i_t,j_1}x_{i_t,j_2}\Big)\mathrm{E}(x_{k,j_1}x_{k,j_2})\prod_{t=m+1}^{a} \mathrm{E}(x_{i_t,j_1}x_{i_t,j_2}) = 0.
\end{aligned}
$$

Therefore, we know (B.65) $\neq 0$ when $k \in \{i_1, \ldots, i_a\}$ and $i_1, \ldots, i_a \leq k$.

When $k < a$, there exist some indexes in $\{i_1, \ldots, i_a\} > k$. Thus (B.65) $= 0$, and (B.64) $= 0$. When $k \geq a$, assume without loss of generality that $i_a = k$ and $i_1, \cdots, i_{a-1} \leq k-1$, then

$$
\mathrm{E}_{k-1}\Big[\Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}\Big)x_{k,j_1}x_{k,j_2}\Big] = \Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}\Big)\mathrm{E}(x_{k,j_1}x_{k,j_2}) = 0,
$$

and

$$
\mathrm{E}_{k}\Big[\Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}\Big)x_{k,j_1}x_{k,j_2}\Big] = \Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}\Big)x_{k,j_1}x_{k,j_2}.
$$

In summary, for $k \geq a$,

$$
\begin{aligned}
&(\mathrm{E}_k - \mathrm{E}_{k-1}) \frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \\
&= \frac{1}{\sigma(a) P^n_a} \sum_{\substack{1 \leq i_1 \neq \cdots \neq i_{a-1} \leq k-1; \\ 1 \leq j_1 \neq j_2 \leq p}} \binom{a}{1} \times (\mathrm{E}_k - \mathrm{E}_{k-1}) \Big[\Big(\prod_{t=1}^{a-1} x_{i_t, j_1} x_{i_t, j_2} \Big) x_{k, j_1} x_{k, j_2} \Big] \\
&= \frac{a}{\sigma(a) P^n_a} \sum_{1 \leq i_1 \neq \cdots \neq i_{a-1} \leq k-1} \sum_{1 \leq j_1 \neq j_2 \leq p} (x_{k,j_1} x_{k,j_2}) \times \prod_{t=1}^{a-1} (x_{i_t,j_1} x_{i_t,j_2}).
\end{aligned}
$$

B.5.6 Proof of Lemma B.1.5

By Lemma B.1.4, we know the explicit form of $D_{n,k} = \sum_{r=1}^{m} t_r A_{n,k,a_r}$, and it follows that $\pi^2_{n,k} = \sum_{1 \leq r_1, r_2 \leq m} t_{r_1} t_{r_2} \mathrm{E}_{k-1}(A_{n,k,a_{r_1}} A_{n,k,a_{r_2}})$. Note that by Cauchy-Schwarz inequality, for some constant C,

$$
\mathrm{var}\Big(\sum_{k=1}^{n} \pi^2_{n,k} \Big) \leq C n^2 \max_{1 \leq k \leq n;\, 1 \leq r_1, r_2 \leq m} \mathrm{var}(\mathbb{T}_{k, a_{r_1}, a_{r_2}}),
$$

where we define $c(n, a_r) = [a_r \times \{\sigma(a_r) P^n_{a_r}\}^{-1}]^2$ and

$$
\begin{aligned}
\mathbb{T}_{k, a_{r_1}, a_{r_2}} &= \mathrm{E}_{k-1}(A_{n,k,a_{r_1}} A_{n,k,a_{r_2}}) \\
&= \sum_{\substack{\mathbf{i} \in \mathcal{P}(k-1, a_{r_1}-1), \\ \tilde{\mathbf{i}} \in \mathcal{P}(k-1, a_{r_2}-1)}} \sum_{\substack{1 \leq j_1 \neq j_2 \leq p, \\ 1 \leq j_3 \neq j_4 \leq p}} \{c(n, a_{r_1}) c(n, a_{r_2})\}^{1/2} \\
&\quad \times \mathrm{E}\Big(\prod_{t=1}^{4} x_{k, j_t} \Big) \times \Big(\prod_{t=1}^{a_{r_1}-1} x_{i_t, j_1} x_{i_t, j_2} \Big) \times \Big(\prod_{t=1}^{a_{r_2}-1} x_{\tilde{i}_t, j_3} x_{\tilde{i}_t, j_4} \Big).
\end{aligned}
$$

Therefore to prove Lemma B.1.5, it suffices to prove $\mathrm{var}(\mathbb{T}_{k, a_{r_1}, a_{r_2}}) = o(n^{-2})$ for every $1 \leq k \leq n$ and $1 \leq r_1, r_2 \leq m$.

Without loss of generality, we prove $\mathrm{var}(\mathbb{T}_{k, a_1, a_2}) = o(n^{-2})$ for any fixed constants a_1 and a_2 and $1 \leq k \leq n$. Similarly to Section B.5.2, for illustration, we first consider a simple setting where $x_{i,j}$'s are independent in Section B.5.6. Next in Section B.5.6.1, we prove that under Condition 3.2.2, $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-2}p^{-1}\log^3 p) = o(n^{-2})$. Last in Section B.5.6.2, we prove that under Condition 3.2.2*, $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-2}p^{-2} + n^{-3}) = o(n^{-2})$. Then Lemma B.1.5 is proved.

Proof illustration In this section, we assume $x_{i,j}$'s are independent and prove $\mathbb{T}_{k,a_1,a_2} = o(n^{-2})$.

When $x_{i,j}$'s are independent, since $j_1 \neq j_2$ and $j_3 \neq j_4$, we know that $\mathrm{E}(x_{k,j_1}x_{k,j_2} \times x_{k,j_3}x_{k,j_4}) \neq 0$ only when $\{j_1, j_2\} = \{j_3, j_4\}$; and it follows that $\mathrm{E}(x_{k,j_1}x_{k,j_2}x_{k,j_3}x_{k,j_4}) = \mathrm{E}(x_{1,j_1}^2)\mathrm{E}(x_{1,j_2}^2)$. Thus $\mathbb{T}_{k,a_1,a_2} = 2c(n,a) \times T_{k,a_1,a_2}$, where we define

$$T_{k,a_1,a_2} = \sum_{\substack{\mathbf{i}\in\mathcal{P}(k-1,a_1-1),\\ \tilde{\mathbf{i}}\in\mathcal{P}(k-1,a_2-1)}} \sum_{1\leq j_1\neq j_2\leq p} \prod_{t=1}^{2}\mathrm{E}(x_{1,j_t}^2)\Big(\prod_{t=1}^{a_1-1} x_{i_t,j_1}x_{i_t,j_2}\Big)\Big(\prod_{t=1}^{a_2-1} x_{\tilde{i}_t,j_1}x_{\tilde{i}_t,j_2}\Big).$$

We note that $c(n,a)$ is of order $\Theta(p^{-2}n^{-a})$ by Lemma B.1.1. To prove $\mathrm{var}(\mathbb{T}_{k,a,a}) = o(n^{-2})$, it suffices to show that $\mathrm{var}(T_{k,a_1,a_2}) = o(n^{a_1+a_2-2}p^4)$. If $a_1 = a_2 = 1$, T_{k,a_1,a_2} is not random and thus $\mathrm{var}(T_{k,a_1,a_2}) = 0$. It remains to consider $a_1 \geq 1$ or $a_2 \geq 1$ below. To examine $\mathrm{var}(T_{k,a_1,a_2})$, we will first consider $\mathrm{E}(T_{k,a_1,a_2})$ and $\mathrm{E}(T_{k,a_1,a_2}^2)$, then $\mathrm{var}(T_{k,a_1,a_2}) = \mathrm{E}(T_{k,a_1,a_2}^2) - \{\mathrm{E}(T_{k,a_1,a_2})\}^2$.

For $\mathrm{E}(T_{k,a_1,a_2})$, note that $\mathrm{E}\{(\prod_{t=1}^{a_1-1} x_{i_t,j_1}x_{i_t,j_2})(\prod_{t=1}^{a_2-1} x_{\tilde{i}_t,j_1}x_{\tilde{i}_t,j_2})\} \neq 0$ only when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ for given $\mathbf{i} \in \mathcal{P}(k-1, a_1-1)$ and $\tilde{\mathbf{i}} \in \mathcal{P}(k-1, a_2-1)$. Therefore, if $a_1 \neq a_2$, $\mathrm{E}(T_{k,a_1,a_2}) = 0$. If $a_1 = a_2 = a$ for some a, we have

$$\mathrm{E}(T_{k,a_1,a_2}) = \sum_{\substack{\mathbf{i}\in\mathcal{P}(k-1,a-1),\\ \tilde{\mathbf{i}}\in\mathcal{P}(k-1,a-1)}} \mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}\}} \sum_{1\leq j_1\neq j_2\leq p} \{\mathrm{E}(x_{1,j_1}^2)\mathrm{E}(x_{1,j_2}^2)\}^a, \tag{B.66}$$

where $\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}\}}$ represents an indicator such that the two sets $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$; and we write

$$\{\mathrm{E}(T_{k,a_1,a_2})\}^2 = \sum_{\substack{\mathbf{i},\mathbf{m}\in\mathcal{P}(k-1,a-1),\\ \tilde{\mathbf{i}},\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a-1)}} \sum_{\substack{1\leq j_1\neq j_2\leq p,\\ 1\leq j_3\neq j_4\leq p}} \mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\,\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}\}} \prod_{t=1}^{4}\{\mathrm{E}(x_{1,j_t}^2)\}^a.$$

where $\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\,\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}\}}$ represents an indicator such that $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\} = \{\tilde{\mathbf{m}}\}$ hold at the same time.

For $\mathrm{E}(T^2_{k,a_1,a_2})$, we have

$$\mathrm{E}(T^2_{k,a_1,a_2}) = \sum_{\substack{\mathbf{i},\,\mathbf{m}\in\mathcal{P}(k-1,a_1-1),\\ \tilde{\mathbf{i}},\,\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a_2-1)}} \sum_{\substack{1\le j_1\ne j_2\le p,\\ 1\le j_3\ne j_4\le p}} \tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}), \tag{B.67}$$

where for the simplicity of notation, we define

$$\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}) = \mathrm{E}\Big(\prod_{t=1}^{a-1} x_{i_t,j_1}x_{i_t,j_2}x_{\tilde{i}_t,j_1}x_{\tilde{i}_t,j_2}x_{m_t,j_3}x_{m_t,j_4}x_{\tilde{m}_t,j_3}x_{\tilde{m}_t,j_4}\Big)\prod_{t=1}^{4}\mathrm{E}(x^2_{1,j_t}).$$

We decompose $\mathrm{E}(T^2_{k,a_1,a_2}) = \mathrm{E}(T^2_{k,a_1,a_2})_{(1)} + \mathrm{E}(T^2_{k,a_1,a_2})_{(2)}$, where

$$\begin{aligned}
\mathrm{E}(T^2_{k,a_1,a_2})_{(1)} &= \sum_{\substack{\mathbf{i},\,\mathbf{m}\in\mathcal{P}(k-1,a_1-1),\\ \tilde{\mathbf{i}},\,\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a_2-1)}} \sum_{\substack{1\le j_1\ne j_2\le p,\\ 1\le j_3\ne j_4\le p}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\\ \{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}}\right\}}\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}),\\
\mathrm{E}(T^2_{k,a_1,a_2})_{(2)} &= \sum_{\substack{\mathbf{i},\,\mathbf{m}\in\mathcal{P}(k-1,a_1-1),\\ \tilde{\mathbf{i}},\,\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a_2-1)}} \sum_{\substack{1\le j_1\ne j_2\le p,\\ 1\le j_3\ne j_4\le p}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}\}\ne\{\tilde{\mathbf{i}}\}\text{ or}\\ \{\mathbf{m}\}\ne\{\tilde{\mathbf{m}}\}}\right\}}\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}),
\end{aligned}$$

where the two indicators $\mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\},\,\{\mathbf{m}\}=\{\tilde{\mathbf{m}}\}\}}$ and $\mathbf{1}_{\{\{\mathbf{i}\}\ne\{\tilde{\mathbf{i}}\}\text{ or }\{\mathbf{m}\}\ne\{\tilde{\mathbf{m}}\}\}}$ represent that $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\} = \{\tilde{\mathbf{m}}\}$ hold at the same time or not, respectively. To prove $\mathrm{var}(T_{k,a_1,a_2}) = o(n^{a_1+a_2-2}p^4)$, since $|\mathrm{var}(T_{k,a_1,a_2})| \le |\mathrm{E}(T^2_{k,a_1,a_2})_{(1)} - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| + |\mathrm{E}(T_{k,a_1,a_2})_{(2)}|$, we show $|\mathrm{E}(T^2_{k,a_1,a_2})_{(1)} - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| = o(n^{2(a-1)}p^4)$ and $\mathrm{E}(T_{k,a_1,a_2})_{(2)} = o(n^{a_1+a_2-2}p^4)$, respectively below.

Part I: $|\mathrm{E}(T^2_{k,a_1,a_2})_{(1)} - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| = o(n^{a_1+a_2-2}p^4)$ By the analysis above, $\mathrm{E}(T_{k,a_1,a_2}) = 0$ if $a_1 \ne a_2$. Also we know $\mathrm{E}(T^2_{k,a_1,a_2})_{(1)} = 0$ if $a_1 \ne a_2$, since $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\} = \{\tilde{\mathbf{m}}\}$ will not happen. Thus it remains to consider $a_1 = a_2 = a$ for some a below. By the forms of $\mathrm{E}(T^2_{k,a_1,a_2})_{(1)}$ and $\{\mathrm{E}(T_{k,a_1,a_2})\}^2$, we consider $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and

$\{\mathbf{m}\} = \{\tilde{\mathbf{m}}\}$. If $\{\mathbf{i}\} \cap \{\mathbf{m}\} = \emptyset$,

$$\tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j}) = \prod_{t=1}^{4} \{\mathrm{E}(x_{1,j_t}^2)\}^a, \tag{B.68}$$

where we use the independence between $x_{i,j}$'s and $j_1 \neq j_2$ and $j_3 \neq j_4$. If $\{j_1, j_2\} \cap \{j_3, j_4\} = \emptyset$, (B.68) also holds similarly by the independence between $x_{i,j}$'s. In summary, when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\} = \{\tilde{\mathbf{m}}\}$, we know that $|\mathrm{E}\{\tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j})\} - \prod_{t=1}^{4} \{\mathrm{E}(x_{1,j_t}^2)\}^a| = 0$, if $\{\mathbf{i}\} \cap \{\mathbf{m}\} = \emptyset$ or $\{j_1, j_2\} \cap \{j_3, j_4\} = \emptyset$. It follows that

$$\begin{aligned} & |\mathrm{E}(T_{k,a_1,a_2}^2)_{(1)} - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| \qquad \text{(B.69)} \\ \leq & \sum_{\substack{\mathbf{i}, \mathbf{m} \in \mathcal{P}(k-1, a_1-1), \\ \tilde{\mathbf{i}}, \tilde{\mathbf{m}} \in \mathcal{P}(k-1, a_2-1)}} \sum_{\substack{1 \leq j_1 \neq j_2 \leq p, \\ 1 \leq j_3 \neq j_4 \leq p}} \mathbf{1}_{\left\{ \substack{\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}, \{\mathbf{m}\} = \{\tilde{\mathbf{m}}\}, \{\mathbf{i}\} \cap \{\mathbf{m}\} \neq \emptyset, \\ \{j_1, j_2\} \cap \{j_3, j_4\} \neq \emptyset} \right\}} \\ & \quad \times \left| \tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j}) - \prod_{t=1}^{4} \{\mathrm{E}(x_{1,j_t}^2)\}^a \right| \\ \leq & \, C n^{a_1 + a_2 - 3} p^{4-1} = o(n^{a_1 + a_2 - 2} p^4), \end{aligned}$$

where we use the boundedness of moments in Condition 3.2.1 and the facts:

$$\sum_{\mathbf{i}, \mathbf{m} \in \mathcal{P}(k-1, a_1-1); \tilde{\mathbf{i}}, \tilde{\mathbf{m}} \in \mathcal{P}(k-1, a_2-1)} \mathbf{1}_{\{\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}, \{\mathbf{m}\} = \{\tilde{\mathbf{m}}\}, \{\mathbf{i}\} \cap \{\mathbf{m}\} \neq \emptyset\}} \leq C n^{a_1 + a_2 - 3},$$

$$\sum_{1 \leq j_1 \neq j_2 \leq p; 1 \leq j_3 \neq j_4 \leq p} \mathbf{1}_{\{\{j_1, j_2\} \cap \{j_3, j_4\} \neq \emptyset\}} \leq C p^{4-1}.$$

Part II: $\mathrm{E}(T_{k,a_1,a_2})_{(2)} = o(n^{a_1+a_2-2} p^4)$ We claim that $\tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j}) = 0$ when $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\} \cup \{\mathbf{m}\} \cup \{\tilde{\mathbf{m}}\}| > a_1 + a_2 - 2$, that is, one of the index only appears once in the four index sets. To see this, we assume, without loss of generality, $i_1 \in \{\mathbf{i}\}$ but $i_1 \notin \{\tilde{\mathbf{i}}\} \cup \{\mathbf{m}\} \cup \{\tilde{\mathbf{m}}\}$, then

$$\tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j}) = \mathrm{E}(x_{i_1,j_1} x_{i_1,j_2}) \times \mathrm{E}(\text{the remaining terms}) = 0. \tag{B.70}$$

Thus when $\tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j}) \neq 0$, the union of the four sets satisfies

$$|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\} \cup \{\mathbf{m}\} \cup \{\tilde{\mathbf{m}}\}| \leq a_1 + a_2 - 2. \tag{B.71}$$

In addition, note that we need to consider $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$ or $\{\mathbf{m}\} \neq \{\tilde{\mathbf{m}}\}$ when analyzing $\mathrm{E}(T^2_{k,a_1,a_2})_{(2)}$. Assume, without loss of generality, that there exists an index $i_1 \in \{\mathbf{i}\}$ but $i_1 \notin \{\tilde{\mathbf{i}}\}$. Similarly to (B.70), we have $\tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j}) \neq 0$ only when $i_1 \in \{\mathbf{m}\} \cup \{\tilde{\mathbf{m}}\}$. If $i_1 \in \{\mathbf{m}\}$ and $i_1 \in \{\tilde{\mathbf{m}}\}$,

$$\tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j}) = \mathrm{E}(x_{1,j_1} x_{1,j_3} x_{1,j_4}) \times \mathrm{E}(\text{all the remaining terms}) = 0,$$

as $j_3 \neq j_4$ and $x_{i,j}$'s are independent; if i_1 is only in one of $\{\mathbf{m}\}$ and $\{\tilde{\mathbf{m}}\}$, for example, $i_1 \in \{\mathbf{m}\}$ but $i_1 \notin \{\tilde{\mathbf{m}}\}$, then

$$\tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j}) = \mathrm{E}(x_{1,j_1} x_{1,j_3}) \times \mathrm{E}(\text{all the remaining terms}),$$

which is nonzero only when $j_1 = j_3$. By analyzing the indexes in $\{\tilde{\mathbf{i}}\}$ symmetrically, we further know $\tilde{Q}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{m}, \tilde{\mathbf{m}}, \mathbf{j}) \neq 0$ only when $\{j_1, j_2\} = \{j_3, j_4\}$. Therefore,

$$|\{j_1, j_2, j_3, j_4\}| = 2. \tag{B.72}$$

Combining (B.71) and (B.72), and by the boundedness of moments in Condition 3.2.1, we have

$$|\mathrm{E}(T^2_{k,a_1,a_2})_{(2)}| = O(n^{a_1+a_2-2} p^2). \tag{B.73}$$

In summary, combining (B.69) and (B.73), we have

$$\begin{aligned}
|\mathrm{var}(T_{k,a_1,a_2})| &= |\mathrm{E}(T_{k,a_1,a_2}^2) - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| \\
&\leq |\mathrm{E}(T_{k,a_1,a_2}^2)_{(1)} - \{\mathrm{E}(T_{k,a_1,a_2})\}^2| + |\mathrm{E}(T_{k,a_1,a_2}^2)_{(2)}| \\
&= O(n^{a_1+a_2-3}p^3) + O(n^{a_1+a_2-2}p^2).
\end{aligned}$$

which is $o(n^{a_1+a_2-2}p^4)$.

B.5.6.1 Proof under Condition 3.2.2

Proof idea Section B.5.6 assumes that $x_{i,j}$'s are independent. In this section, we further prove Lemma B.1.5 under Condition 3.2.2. Similarly to Section B.5.2, we know that under Condition 3.2.2, $x_{i,j}$'s may be no longer independent, but the dependence between x_{i,j_1} and x_{i,j_2} degenerates exponentially with their distance $|j_1 - j_2|$. To quantitatively examine $|j_1 - j_2|$, we will introduce a threshold of distance D_0 to be defined in (B.77) below, which is similar to K_0 in (B.40). Intuitively, when $|j_1 - j_2| > D_0$, x_{i,j_1} and x_{i,j_2} are "asymptotically independent" with similar properties to those under the independence case in Section B.5.6. The following proof will provide comprehensive discussions based on D_0.

Recall that as argued at the beginning of Section B.5.6, to prove Lemma B.1.5, it suffices to show $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-2}p^{-1}\log^3 p) = o(n^{-2})$ for any fixed integers a_1 and a_2. To facilitate the discussion, we define some notation to be used in the proof.

Notation For given tuples $\mathbf{i}^{(l)} = (i_1, \dots, i_{a_l-1}) \in \mathcal{P}(k-1, a_l - 1)$ with $l = 1, 2$, we define $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) = (i_1^{(1)}, \dots, i_{a_1-1}^{(1)}, i_1^{(2)}, \dots, i_{a_2-1}^{(2)})$, and let $S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)})$ be a collection of tuples $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)})$ where $\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1)$ for $l = 1, 2$. Moreover, we define

$\mathcal{J} = \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p)$. Then

$$\mathbb{T}_{k,a_1,a_2} = \sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)});\\ (j_1,j_2),(j_3,j_4)\in\mathcal{J}}} \Big\{\prod_{l=1}^{2} c(n,a_l)\Big\}^{1/2} \times \mathbb{X}(k,\mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2),$$

where we recall that $c(n,a) = [a \times \{\sigma(a)P_a^n\}^{-1}]^2$ and we define

$$\mathbb{X}(k,\mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2) = \mathrm{E}\Big(\prod_{t=1}^{4} x_{k,j_t}\Big)\prod_{l=1}^{2}\prod_{t=1}^{a_l-1}(x_{i_t^{(l)},j_{2l-1}} x_{i_t^{(l)},j_{2l}}).$$

In addition, for easy representation, we define $a_3 = a_1$ and $a_4 = a_2$. Then for given tuples $\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1)$ with $l = 1,2,3,4$, we define the tuple

$$(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)}) = (i_1^{(1)},\ldots,i_{a_1-1}^{(1)}, i_1^{(2)},\ldots,i_{a_2-1}^{(2)}, i_1^{(3)},\ldots,i_{a_1-1}^{(3)}, i_1^{(4)},\ldots,i_{a_2-1}^{(4)}),$$

and let $S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)})$ be a collection of $(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)})$ where $\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1)$ with $l = 1,2,3,4$. Then we can write

$$\mathbb{T}_{k,a_1,a_2}^2 = \sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)});\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \prod_{l=1}^{2} c(n,a_l)\,\mathbb{X}(k,\mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4),$$

where we define

$$\begin{aligned}&\mathbb{X}(k,\mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4)\\ &= \mathrm{E}\Big(\prod_{t=1}^{4} x_{k,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8} x_{k,j_t}\Big)\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_{2l-1}} x_{i_t^{(l)},j_{2l}}.\end{aligned}$$

Recall the definitions at the beginning of Section B.5.1. $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$ represents that the two tuples have the same elements without order. We next decompose $S(\mathbf{i}^{(1)},\mathbf{i}^{(2)})$ into two parts: the collection $S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},1)$ contains the tuples $(\mathbf{i}^{(1)},\mathbf{i}^{(2)})$ satisfying $\{\mathbf{i}^{(1)}\} \neq \{\mathbf{i}^{(2)}\}$, and the collection $S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},2)$ contains the tuples $(\mathbf{i}^{(1)},\mathbf{i}^{(2)})$

satisfying $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$. Then we can write $\mathbb{T}_{k,a_1,a_2} = \sum_{v=1}^{2} \mathbb{T}_{k,a_1,a_2,v}$, where

$$\mathbb{T}_{k,a_1,a_2,v} = \sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},v);\\ (j_1,j_2),(j_3,j_4)\in\mathcal{J}}} \Big\{\prod_{l=1}^{2} c(n,a_l)\Big\}^{1/2} \times \mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2).$$

In addition, for $v = 1,2$, we let the collection $S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},v,v)$ contain the tuples $(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)})$ such that $(\mathbf{i}^{(1)},\mathbf{i}^{(2)}) \in S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},v)$ and $(\mathbf{i}^{(3)},\mathbf{i}^{(4)}) \in S(\mathbf{i}^{(3)},\mathbf{i}^{(4)},v)$. It follows that for $v = 1,2$, we can write

$$\begin{aligned}\mathbb{T}^2_{k,a_1,a_2,v} = & \sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},v,v);\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \prod_{l=1}^{2} c(n,a_l) \\ & \times\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4).\end{aligned} \tag{B.74}$$

We next define some notation on the j indexes. Given a tuple $(j_{t_1},j_{t_2},j_{t_3},j_{t_4})$, we write its corresponding ordered version as

$$(\tilde{j}_{t_1},\tilde{j}_{t_2},\tilde{j}_{t_3},\tilde{j}_{t_4}) \quad \text{satisfying} \quad \tilde{j}_{t_1} \le \tilde{j}_{t_2} \le \tilde{j}_{t_3} \le \tilde{j}_{t_4}. \tag{B.75}$$

Given the ordered indexes, we define the maximum distance between indexes in the given tuple as $\mathbb{D}_M(j_{t_1},j_{t_2},j_{t_3},j_{t_4}) = \max\{\tilde{j}_{t_2}-\tilde{j}_{t_1},\tilde{j}_{t_3}-\tilde{j}_{t_2},\tilde{j}_{t_4}-\tilde{j}_{t_3}\}$. For the simplicity of presentation later, for tuples $(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}$, we further define

$$\begin{aligned}\kappa_1 &= \mathbb{D}_M(j_1,j_2,j_3,j_4), & \kappa_2 &= \mathbb{D}_M(j_5,j_6,j_7,j_8),\\ \kappa_3 &= \mathbb{D}_M(j_1,j_2,j_5,j_6) & \kappa_4 &= \mathbb{D}_M(j_1,j_2,j_7,j_8).\end{aligned} \tag{B.76}$$

In the following discussion, to quantitatively evaluate the distances in (B.76), we introduce a threshold D_0 below. In particular, given small positive constants μ and

ϵ, and δ in Condition 3.2.2, we define

$$D_0 = \frac{-(2+\epsilon)(8+\mu)\log p}{\epsilon \log \delta}, \tag{B.77}$$

which will be used as discussed at the beginning of this section on Page 313.

Proof We present the proof of $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-2}p^{-1}\log^3 p)$ based on the notation above. Note that we can write $\mathbb{T}_{k,a_1,a_2} = \sum_{v=1}^{2} \mathbb{T}_{k,a_1,a_2,v}$. By the Cauchy-Schwarz inequality, we know it suffices to show $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,v}) = O(n^{-2}p^{-1}\log^3 p)$ for $v = 1,2$ respectively.

Step I: $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,1}) = O(n^{-2}p^{-1}\log^3 p)$ By the definition of $\mathbb{T}_{k,a_1,a_2,1}$, we have $\{\mathbf{i}^{(1)}\} \neq \{\mathbf{i}^{(2)}\}$ for $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) \in S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, 1)$. Suppose, without loss of generality, that index $i \in \{\mathbf{i}^{(1)}\}$ but $i \notin \{\mathbf{i}^{(2)}\}$. Then under H_0,

$$\mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4)\} = \mathrm{E}(x_{i,j_1}x_{i,j_2}) \times \mathrm{E}(\text{other terms}) = 0. \tag{B.78}$$

Therefore $\mathrm{E}(\mathbb{T}_{k,a_1,a_2,1}) = 0$ and $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,1}) = \mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,1})$.

By (B.74), we have

$$\mathbb{T}^2_{k,a_1,a_2,1} = \sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},1,1);\\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \prod_{l=1}^{2} c(n,a_l) \times \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4).$$

To prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,1}) = O(n^{-2}p^{-1}\log^3 p)$, we will next show that for given indexes $(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in \mathcal{J}$,

$$\mathrm{E}\Big\{\sum_{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},1,1)} \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4)\Big\} = O(n^{a_1+a_2-2}); \tag{B.79}$$

and for given $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}) \in S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, 1, 1)$,

$$\mathrm{E}\Big\{ \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4) \Big\} = O(p^3 \log^3 p). \tag{B.80}$$

Given (B.79) and (B.80), since $c(n, a_l) = \Theta(p^{-2} n^{-a_l})$, we can obtain $\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,1}) = O(n^{-2}p^{-1}\log^3 p)$. Thus to finish the proof, it remains to prove (B.79) and (B.80).

To prove (B.79), we claim that $\mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4)\} = 0$ when $|\cup_{l=1}^4 \{\mathbf{i}^{(l)}\}| > a_1 + a_2 - 2$, i.e., there exists one index only appears once in the four index sets $\{\mathbf{i}^{(l)}\}$, $l = 1, \ldots, 4$. Too see this, suppose an index $i \in \{\mathbf{i}^{(1)}\}$ but $i \notin \{\mathbf{i}^{(2)}\}$, $i \notin \{\mathbf{i}^{(3)}\}$ and $i \notin \{\mathbf{i}^{(4)}\}$, then (B.78) holds. Therefore, $\mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4)\} \neq 0$ only when

$$\Big| \cup_{l=1}^4 \{\mathbf{i}^{(l)}\} \Big| \leq a_1 + a_2 - 2. \tag{B.81}$$

By the boundedness of moments from Condition 3.2.1, we know (B.79) holds.

We next prove (B.80). For $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}) \in S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, 1, 1)$, we know $\{\mathbf{i}^{(1)}\} \neq \{\mathbf{i}^{(2)}\}$ and $\{\mathbf{i}^{(3)}\} \neq \{\mathbf{i}^{(4)}\}$. Suppose, without loss of generality, there exists an index $i \in \{\mathbf{i}^{(3)}\}$ and $i \notin \{\mathbf{i}^{(4)}\}$. If $i \notin \{\mathbf{i}^{(1)}\}$ and $i \notin \{\mathbf{i}^{(2)}\}$, similarly, (B.78) holds. Then we consider $i \in \{\mathbf{i}^{(1)}\}$ or $i \in \{\mathbf{i}^{(2)}\}$ in the following three cases.

Case 1: When $i \in \{\mathbf{i}^{(1)}\}$ and $i \notin \{\mathbf{i}^{(2)}\}$, we know

$$\begin{aligned} & \mathrm{E}\Big\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4)\Big\} \qquad \text{(B.82)} \\ = \ & \mathrm{E}\Big(\prod_{t=1}^4 x_{k,j_t}\Big) \times \mathrm{E}\Big(\prod_{t=5}^8 x_{k,j_t}\Big) \times \mathrm{E}\Big(\prod_{t=1,2,5,6} x_{i,j_t}\Big) \times \mathrm{E}(\text{other terms}). \end{aligned}$$

If $x_{i,j}$'s are independent as in Section B.5.6, we know (B.82) $\neq 0$ only when $\{j_1, j_2\} =$

$\{j_3, j_4\} = \{j_5, j_6\} = \{j_7, j_8\}$, which induces $|\{j_1, \ldots, j_8\}| = 2$ and

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}} \mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\} = O(p^2),$$

i.e., (B.80) is obtained. Under Condition 3.2.2, $x_{i,j}$'s may be no longer independent, but as discussed at the beginning of Section B.5.6.1, we can still prove (B.80) similarly to the independence case. In particular, based on D_0 in (B.77), we evaluate (B.82) by discussing the following three sub-cases (a)–(c).

(a) When both (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) contain only two distinct indexes within each tuple, i.e., $|\{j_1, j_2, j_3, j_4\}| = |\{j_5, j_6, j_7, j_8\}| = 2$, we consider without loss of generality that $j_1 = j_3$, $j_2 = j_4$, $j_5 = j_7$, and $j_6 = j_8$. Then

$$(\text{B.82}) = \mathrm{E}(x_{k,j_1}^2 x_{k,j_2}^2)\mathrm{E}(x_{k,j_5}^2 x_{k,j_6}^2)\mathrm{E}(x_{k,j_1} x_{k,j_2} x_{k,j_5} x_{k,j_6})\mathrm{E}(\text{other terms}).$$

(a.1) If (j_1, j_2, j_5, j_6) contains two distinct indexes, i.e., $|\{j_1, j_2, j_5, j_6\}| = 2$, we assume without loss of generality that $j_1 = j_5$ and $j_2 = j_6$. Then $|\{j_1, \ldots, j_8\}| = 2$ and in this case, the total number of distinct j indexes is $O(p^2)$.

(a.2) If (j_1, j_2, j_5, j_6) contains at least three distinct indexes, that is, $|\{j_1, j_2, j_5, j_6\}| \geq 3$, we have $|\{\tilde{j}_1, \tilde{j}_2, \tilde{j}_5, \tilde{j}_6\}| \geq 3$, where $(\tilde{j}_1, \tilde{j}_2, \tilde{j}_5, \tilde{j}_6)$ denotes the ordered version of (j_1, j_2, j_5, j_6) following the notation in (B.75). Then we have $\mathrm{E}(x_{k,\tilde{j}_1} x_{k,\tilde{j}_2}) \times \mathrm{E}(x_{k,\tilde{j}_5} x_{k,\tilde{j}_6}) = 0$. Together with $\mathrm{E}(\mathbf{x}) = \mathbf{0}$, we can write

$$\begin{aligned} |\mathrm{E}(x_{1,j_1} x_{1,j_2} x_{1,j_5} x_{1,j_6})| &= |\mathrm{cov}(x_{k,\tilde{j}_1} x_{k,\tilde{j}_2}\ ,\ x_{k,\tilde{j}_5} x_{k,\tilde{j}_6})| \\ &= |\mathrm{cov}(x_{k,\tilde{j}_1}\ ,\ x_{k,\tilde{j}_2} x_{k,\tilde{j}_5} x_{k,\tilde{j}_6})| \\ &= |\mathrm{cov}(x_{k,\tilde{j}_1} x_{k,\tilde{j}_2} x_{k,\tilde{j}_5}\ ,\ x_{k,\tilde{j}_6})|. \end{aligned} \tag{B.83}$$

Recall that κ_3 in (B.76) represents the maximum distance between (j_1, j_2, j_5, j_6).

If $\kappa_3 > D_0$, by Conditions 3.2.1 and 3.2.2, and the α-mixing inequality in Lemma B.5.1, we know

$$|(\text{B.82})| \le C \times (\text{B.83}) \le C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}).$$

If $\kappa_3 \le D_0$, the total number of distinct j indexes is $O(pD_0^3)$.

(b) When both (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) have at least 3 distinct elements, i.e., $|\{j_1, j_2, j_3, j_4\}| \ge 3$ and $|\{j_5, j_6, j_7, j_8\}| \ge 3$, following the notation in (B.75), similarly to (B.83), we can write

$$\begin{aligned} |\mathrm{E}(x_{k,j_1}x_{k,j_2}x_{k,j_3}x_{k,j_4})| &= |\mathrm{cov}(x_{k,\tilde{j}_1}x_{k,\tilde{j}_2}\ ,\ x_{k,\tilde{j}_3}x_{k,\tilde{j}_4})| \\ &= |\mathrm{cov}(x_{k,\tilde{j}_1}\ ,\ x_{k,\tilde{j}_2}x_{k,\tilde{j}_3}x_{k,\tilde{j}_4})| \\ &= |\mathrm{cov}(x_{k,\tilde{j}_1}x_{k,\tilde{j}_2}x_{k,\tilde{j}_3}\ ,\ x_{k,\tilde{j}_4})|, \end{aligned} \tag{B.84}$$

and

$$\begin{aligned} |\mathrm{E}(x_{k,j_5}x_{k,j_6}x_{k,j_7}x_{k,j_8})| &= |\mathrm{cov}(x_{k,\tilde{j}_5}x_{k,\tilde{j}_6}\ ,\ x_{k,\tilde{j}_7}x_{k,\tilde{j}_8})| \\ &= |\mathrm{cov}(x_{k,\tilde{j}_5}\ ,\ x_{k,\tilde{j}_6}x_{k,\tilde{j}_7}x_{k,\tilde{j}_8})| \\ &= |\mathrm{cov}(x_{k,\tilde{j}_5}x_{k,\tilde{j}_6}x_{k,\tilde{j}_7}\ ,\ x_{k,\tilde{j}_8})|. \end{aligned} \tag{B.85}$$

When $\max\{\kappa_1, \kappa_2\} > D_0$ in this case, by Conditions 3.2.1 and 3.2.2, and the α-mixing inequality,

$$|(\text{B.82})| \le C \times (\text{B.84}) \times (\text{B.85}) \le C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}). \tag{B.86}$$

When $\max\{\kappa_1, \kappa_2\} \le D_0$, by the definitions in (B.76), we know under this case, the indexes in (j_1, j_2, j_3, j_4) are close to each other within the distance D_0, and the indexes in (j_5, j_6, j_7, j_8) are also close to each other within the distance D_0. Then the total number of distinct indexes is $O(pD_0^3 \times pD_0^3) = O(p^2 D_0^6)$.

(c) If only one of (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) contains at least 3 distinct indexes, without loss of generality, we assume $|\{j_1, j_2, j_3, j_4\}| \geq 3$ and $|\{j_5, j_6, j_7, j_8\}| = 2$. When $\kappa_1 \leq D_0$, the indexes in (j_1, j_2, j_3, j_4) are close within distance D_0. As (j_5, j_6, j_7, j_8) only contains 2 distinct indexes, the total number of distinct j indexes is $O(p^3 D_0^3)$. When $\kappa_1 > D_0$, by Conditions 3.2.1 and 3.2.2, and the α-mixing inequality, we know

$$|(\text{B.82})| \leq C \times (\text{B.84}) \leq C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}). \tag{B.87}$$

Case 2: When $i \notin \{\mathbf{i}^{(1)}\}$ and $i \in \{\mathbf{i}^{(2)}\}$, we know similar conclusion holds by symmetricity.

Case 3: When $i \in \{\mathbf{i}^{(1)}\}$ and $i \in \{\mathbf{i}^{(2)}\}$, we have

$$\begin{aligned} & \mathrm{E}\Big\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2,3,4)\Big\} \\ = \;& \mathrm{E}\Big(\prod_{t=1}^{4} x_{k,j_t}\Big) \times \mathrm{E}\Big(\prod_{t=5}^{8} x_{k,j_t}\Big) \times \mathrm{E}\Big(\prod_{t=1}^{6} x_{k,j_t}\Big) \times \mathrm{E}(\text{other terms}) \end{aligned} \tag{B.88}$$

Similarly to **Case 1** above, to evaluate (B.88), we next discuss two sub-cases with D_0 in (B.77).

(a) When both (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) only contain 2 distinct indexes within each tuple, i.e., $|\{j_1, j_2, j_3, j_4\}| = |\{j_5, j_6, j_7, j_8\}| = 2$, we assume $j_1 = j_3$, $j_2 = j_4$, $j_5 = j_7$ and $j_6 = j_8$ without loss of generality. Then

$$(\text{B.88}) = \mathrm{E}(x_{k,j_1}^2 x_{k,j_2}^2)\mathrm{E}(x_{k,j_5}^2 x_{k,j_6}^2)\mathrm{E}(x_{i,j_1}^2 x_{i,j_2}^2 x_{i,j_5} x_{i,j_6})\mathrm{E}(\text{other terms}).$$

Following the notation in (B.75), when $\tilde{k}_3^* := \min\{\tilde{j}_2 - \tilde{j}_1, \tilde{j}_5 - \tilde{j}_2, \tilde{j}_6 - \tilde{j}_5\} < D_0$, the total number of distinct j indexes is $O(p^3 D_0)$. When $\tilde{k}_3^* > D_0$, by Conditions 3.2.1, 3.2.2, and the α-mixing inequality,

$$
\begin{aligned}
&|\mathrm{E}(x_{1,\tilde{j}_1}^2 x_{1,\tilde{j}_2}^2 x_{1,\tilde{j}_5} x_{1,\tilde{j}_6})| \\
=& |\mathrm{cov}(x_{1,\tilde{j}_1}^2 \ , \ x_{1,\tilde{j}_2}^2 x_{1,\tilde{j}_5} x_{1,\tilde{j}_6}) + \mathrm{E}(x_{1,\tilde{j}_1}^2)\mathrm{cov}(x_{1,\tilde{j}_2}^2 \ , \ x_{1,\tilde{j}_5} x_{1,\tilde{j}_6}) \\
&+ [\mathrm{E}(x_{1,\tilde{j}_1}^2)]^2 \mathrm{cov}(x_{1,\tilde{j}_5} \ , \ x_{1,\tilde{j}_6})| \\
\leq& C\delta^{\frac{D_0 \epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}).
\end{aligned}
$$

(b) If at least one of (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) has at least 3 distinct indexes within the tuple, it means that $|\{j_1, j_2, j_3, j_4\}| \geq 3$ or $|\{j_5, j_6, j_7, j_8\}| \geq 3$. Similarly to (B.86) and (B.87), we know that when $\max\{\kappa_1, \kappa_2\} > D_0$, $|(\text{B.88})| = O(p^{-(8+\mu)})$; when $\max\{\kappa_1, \kappa_2\} \leq D_0$, the total number of distinct j indexes is $O(p^3 D_0^3)$.

Combining Cases 1–3 discussed above, we obtain

$$
\begin{aligned}
&\mathrm{E}\Big\{ \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4) \Big\} \\
=& O(p^3 D_0^3) + \sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}} O(p^{-(8+\mu)}) \\
=& O(p^3 \log^3 p) + p^8 O(p^{-(8+\mu)}) = O(p^3 \log^3 p),
\end{aligned}
$$

where we use $\mu > 0$ and $D_0 = O(\log p)$ by (B.77). Thus (B.80) is proved.

Step II: $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,2}) = O(n^{-2}p^{-1}\log^3 p)$ Recall that $\mathbb{T}_{k,a_1,a_2,2}$ is constructed from $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}) \in S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, 2)$, where $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$. As $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$ happens only when $a_1 = a_2$, so it remains to consider $a_1 = a_2 = a$ for some integer a below. It follows that $\mathrm{E}\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\}=\{\mathrm{E}(\prod_{t=1}^{4}x_{1,j_t})\}^a$, and then

$$\mathrm{E}(\mathbb{T}_{k,a_1,a_2,2})=\sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},2);\\(j_1,j_2),(j_3,j_4)\in\mathcal{J}}}\Big\{\prod_{l=1}^{2}c(n,a_l)\Big\}^{1/2}\times\Big\{\mathrm{E}\Big(\prod_{t=1}^{4}x_{1,j_t}\Big)\Big\}^a,$$

and

$$\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,2})\}^2=\sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2);\\(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}}}\prod_{l=1}^{2}c(n,a_l)\Big\{\mathrm{E}\Big(\prod_{t=1}^{4}x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8}x_{1,j_t}\Big)\Big\}^a.$$

Moreover, by (B.74), we know $\mathbb{T}^2_{k,a_1,a_2,2}$ is a summation over $(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{4)})\in S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2)$, where $\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}$ and $\{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}$ by the construction. We define $S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2,q)$ to be the collection of tuples $(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{4)})$ such that $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}|=q$, where $0\le q\le a-1$. Then we write $\mathbb{T}^2_{k,a_1,a_2,2}=\sum_{q=0}^{a-1}\mathbb{T}^2_{k,a_1,a_2,2,(q)}$, where we define

$$\mathbb{T}^2_{k,a_1,a_2,2,(q)}=\sum_{\substack{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2,q);\\(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}}}\prod_{l=1}^{2}c(n,a_l)\times\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4).$$

In particular, when $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}|=q$,

$$\mathrm{E}\Big\{\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4)\Big\}=\Big\{\mathrm{E}\Big(\prod_{t=1}^{4}x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8}x_{1,j_t}\Big)\Big\}^{a-q}\Big\{\prod_{t=1}^{8}x_{1,j_t}\Big\}^q.$$

Therefore, for $a_1=a_2=a$,

$$\begin{aligned}\mathrm{var}(\mathbb{T}_{k,a_1,a_2,2}) &= \mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,2})-\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,2})\}^2\\ &= \sum_{q=0}^{a-1}\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,2,(q)})-\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,2})\}^2\\ &= \sum_{q=1}^{a-1}\sum_{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2,q)}\prod_{l=1}^{2}c(n,a_l)\times\mathbb{D}_{k,a,a,2,q},\end{aligned}$$

where we define

$$\mathbb{D}_{k,a,a,2,q} = \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8} x_{1,j_t}\Big)\Big\}^{a-q} \times\Bigg[\Big\{\mathrm{E}\Big(\prod_{t=1}^{8} x_{1,j_t}\Big)\Big\}^{q} - \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8} x_{1,j_t}\Big)\Big\}^{q}\Bigg],$$

and use $\mathbb{D}_{k,a,a,2,q} = 0$ when $q = 0$. By the construction, we know the total number of tuples in the collection $S(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, 2, 2, q)$ is bounded by $Cn^{2(a-1)-q}$, that is, for some constant C,

$$\sum_{S(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},2,2,q)} 1 \le Cn^{2(a-1)-q}. \tag{B.89}$$

Since $c(n,a) = \Theta(p^{-2}n^{-a})$, to prove $\mathrm{var}(\mathbb{T}^2_{k,a_1,a_2,2}) = O(n^{-2}p^{-1}\log^3 p)$, it suffices to show for given tuple $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)})$, $\mathbb{D}_{k,a_1,a_2,2,q} = O(p^3 \log^3 p)$ for $1 \le q \le a-1$.

By Condition 3.2.1 and Lemma B.5.2 (on Page 286), for $1 \le q \le a-1$,

$$|\mathbb{D}_{k,a,a,2,q}| \le C \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \left|\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\right| \times \left|\mathrm{E}\Big(\prod_{t=5}^{8} x_{1,j_t}\Big)\right| \times\left|\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big) \times \mathrm{E}\Big(\prod_{t=5}^{8} x_{1,j_t}\Big) - \mathrm{E}\Big(\prod_{t=1}^{8} x_{1,j_t}\Big)\right|.$$

To evaluate $\mathbb{D}_{k,a,a,2,q}$, we next discuss several cases, based on the notation $\kappa_1, \ldots, \kappa_4$ in (B.76), and D_0 in (B.77).

(a) When both tuples (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) contain only two distinct indexes, i.e., $|\{j_1, j_2, j_3, j_4\}| = |\{j_5, j_6, j_7, j_8\}| = 2$, we assume without loss of generality that $j_1 = j_3$, $j_2 = j_4$, $j_5 = j_7$ and $j_6 = j_8$. Then $\mathrm{E}(\prod_{t=1}^{4} x_{1,j_t}) = \mathrm{E}(x^2_{1,j_1}x^2_{1,j_2})$, $\mathrm{E}(\prod_{t=5}^{8} x_{1,j_t}) = \mathrm{E}(x^2_{1,j_5}x^2_{1,j_6})$ and $\mathrm{E}(\prod_{t=1}^{8} x_{1,j_t}) = \mathrm{E}(x^2_{1,j_1}x^2_{1,j_2}x^2_{1,j_5}x^2_{1,j_6})$.

Following the notation in (B.75), let $(\tilde{j}_1 \le \tilde{j}_2 \le \tilde{j}_5 \le \tilde{j}_6)$ be the ordered version of (j_1, j_2, j_5, j_6). When $\min\{\tilde{j}_2 - \tilde{j}_1, \tilde{j}_5 - \tilde{j}_2, \tilde{j}_6 - \tilde{j}_5\} \le D_0$, the total number of distinct j indexes is $O(p^3 D_0)$. When $\min\{\tilde{j}_2 - \tilde{j}_1, \tilde{j}_5 - \tilde{j}_2, \tilde{j}_6 - \tilde{j}_5\} > D_0$, by Conditions 3.2.1 and 3.2.2, and the α-mixing inequality in Lemma B.5.1,

$$
\begin{aligned}
&\Big|\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8} x_{1,j_t}\Big) - \mathrm{E}\Big(\prod_{t=1}^{8} x_{1,j_t}\Big)\Big| \\
=& \Big|\mathrm{E}(x_{1,j_1}^2 x_{1,j_2}^2)\mathrm{E}(x_{1,j_5}^2 x_{1,j_6}^2) - \mathrm{E}(x_{1,j_1}^2 x_{1,j_2}^2 x_{1,j_5}^2 x_{1,j_6}^2)\Big| \\
\le& C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}).
\end{aligned}
$$

(b) When both (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) contain at least 3 distinct indexes, i.e., $|\{j_1, j_2, j_3, j_4\}| \ge 3$ and $|\{j_5, j_6, j_7, j_8\}| \ge 3$, we know similarly (B.84) and (B.85) hold. When $\max\{\kappa_1, \kappa_2\} > D_0$, by Conditions 3.2.1 and 3.2.2, and the α-mixing inequality in Lemma B.5.1, we obtain

$$
|\mathbb{D}_{k,a_1,a_2,2,q}| \le C(\text{B.84}) \times (\text{B.85}) \le C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O\{p^{-(8+\mu)}\}.
$$

When $\max\{\kappa_1, \kappa_2\} \le D_0$, by the definitions in (B.76), we know under this case, the indexes in (j_1, j_2, j_3, j_4) are close to each other within the distance D_0, and the indexes in (j_5, j_6, j_7, j_8) are also close to each other within the distance D_0. Then the total number of distinct j indexes is $O(pD_0^3 \times pD_0^3) = O(p^2 D_0^6)$.

(c) When only one of (j_1, j_2, j_3, j_4) and (j_5, j_6, j_7, j_8) contains at least 3 distinct indexes, without loss of generality, assume $|\{j_1, j_2, j_3, j_4\}| \ge 3$ and $|\{j_5, j_6, j_7, j_8\}| = 2$. Recall κ_1 defined in (B.76). When $\kappa_1 \le D_0$, the indexes in (j_1, j_2, j_3, j_4) are close within distance D_0. As (j_5, j_6, j_7, j_8) only contains 2 distinct indexes, the total number of distinct j indexes is $O(p^3 D_0^3)$. When $\kappa_1 > D_0$, by Conditions 3.2.1 and 3.2.2, and the α-mixing inequality in Lemma B.5.1, we know similarly

(B.84) holds, and

$$|\mathbb{D}_{k,a_1,a_2,2,q}| \le C(\text{B.84}) \le C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}).$$

In summary,

$$|\mathbb{D}_{k,a_1,a_2,2,q}| = p^8 \times O(p^{-(8+\mu)}) + O(p^3 D_0^3) = O(p^3 \log^3 p). \tag{B.90}$$

Thus we obtain that for given $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)})$, $\mathbb{D}_{k,a_1,a_2,2,q} = O(p^3 \log^3 p)$. Combined with (B.89), $\text{var}(\mathbb{T}^2_{k,a_1,a_2,2}) = O(n^{-2}p^{-1}\log^3 p)$ follows.

Combining the results in *Step I* and *Step II* above, we obtain $\text{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-2}p^{-1}\log^3 p)$, and thus Lemma B.1.5 is proved under Condition 3.2.2.

B.5.6.2 Proof under Condition 3.2.2*

In this section, we prove Lemma B.1.5 by substituting Condition 3.2.2 with Condition 3.2.2*. Note that although the independence between $x_{i,j}$'s is assumed in Section B.5.6, it is only used to specify certain joint moments of $x_{i,j}$'s. Alternatively, Condition 3.2.2* is assumed to obtain similar properties on the joint moments, and the proof follows similarly to that in Section B.5.6.

In particular, we will prove that $\text{var}(\mathbb{T}_{k,a_1,a_2}) = O(n^{-3} + n^{-2}p^{-2})$ for two given finite integers a_1 and a_2 below. Under H_0 and given Condition 3.2.2*, as $j_1 \neq j_2$ and $j_3 \neq j_4$, we have $\text{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) \neq 0$ only when $\{j_1, j_2\} = \{j_3, j_4\}$, and then $\text{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) = \kappa_1 \text{E}(x^2_{1,j_1})\text{E}(x^2_{1,j_2})$. It follows that $\mathbb{T}_{k,a_1,a_2} = 2c(n,a) \times \tilde{T}_{k,a_1,a_2}$, where $\tilde{T}_{k,a,a} = \kappa_1 T_{k,a,a}$ with $T_{k,a,a}$ defined in Section B.5.6. To prove $\text{var}(\mathbb{T}_{k,a,a}) = o(n^{-2})$, it suffices to show that $\text{var}(\tilde{T}_{k,a_1,a_2}) = n^{a_1+a_2-2}p^4 O(n^{-1} + p^{-2})$ as argued in Section B.5.6.

Similarly to Section B.5.6, to show $\text{var}(\tilde{T}_{k,a_1,a_2}) = n^{a_1+a_2-2}p^4 O(n^{-1} + p^{-2})$, we examine $\{\text{E}(\tilde{T}_{k,a_1,a_2})\}^2$ and $\text{E}(\tilde{T}^2_{k,a_1,a_2})$ respectively. For $\text{E}(\tilde{T}_{k,a_1,a_2})$, under Condition

3.2.2*, similarly to (B.66), we know $\mathrm{E}\{(\prod_{t=1}^{a_1-1} x_{i_t,j_1}x_{i_t,j_2}) \times (\prod_{t=1}^{a_2-1} x_{\tilde{i}_t,j_1}x_{\tilde{i}_t,j_2})\} \neq 0$ only when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$. When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$, we write $a_1 = a_2 = a$ for some a and then $\mathrm{E}\{(\prod_{t=1}^{a_1-1} x_{i_t,j_1}x_{i_t,j_2}) \times (\prod_{t=1}^{a_2-1} x_{\tilde{i}_t,j_1}x_{\tilde{i}_t,j_2})\} = \{\kappa_1 \mathrm{E}(x_{1,j_1}^2)\mathrm{E}(x_{1,j_2}^2)\}^{a-1}$. We thus have $\{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2 = \{\kappa_1^a \mathrm{E}(T_{k,a_1,a_2})\}^2$ with T_{k,a_1,a_2} defined in Section B.5.6. Moreover, following (B.67) in Section B.5.6, we have

$$\mathrm{E}(\tilde{T}_{k,a_1,a_2}^2) = \sum_{\substack{\mathbf{i},\,\mathbf{m}\in\mathcal{P}(k-1,a_1-1);\\ \tilde{\mathbf{i}},\,\tilde{\mathbf{m}}\in\mathcal{P}(k-1,a_2-1)}} \sum_{\substack{1\le j_1\neq j_2\le p;\\ 1\le j_3\neq j_4\le p}} \tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j}).$$

We further decompose $\mathrm{E}(\tilde{T}_{k,a_1,a_2}^2) = \mathrm{E}(\tilde{T}_{k,a_1,a_2}^2)_{(1)} + \mathrm{E}(\tilde{T}_{k,a_1,a_2}^2)_{(2)}$, where $\mathrm{E}(\tilde{T}_{k,a_1,a_2}^2)_{(1)}$ and $\mathrm{E}(\tilde{T}_{k,a_1,a_2}^2)_{(2)}$ are defined with the same forms as $\mathrm{E}(T_{k,a_1,a_2}^2)_{(1)}$ and $\mathrm{E}(T_{k,a_1,a_2}^2)_{(2)}$ in Section B.5.6, respectively. To prove $\mathrm{var}(\tilde{T}_{k,a_1,a_2}) = n^{a_1+a_2-2}p^4 O(n^{-1}+p^{-2})$, similarly to Section B.5.6, we derive $|\mathrm{E}(\tilde{T}_{k,a_1,a_2}^2)_{(1)} - \{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2|$ and $\mathrm{E}(T_{k,a_1,a_2}^2)_{(2)}$ respectively.

Step I: $|\mathrm{E}(\tilde{T}_{k,a_1,a_2}^2)_{(1)} - \{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2|$ By the forms of $\mathrm{E}(\tilde{T}_{k,a_1,a_2}^2)_{(1)}$ and $\mathrm{E}(\tilde{T}_{k,a_1,a_2})$, we consider $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{m}\} = \{\tilde{\mathbf{m}}\}$ below. If $\{\mathbf{i}\}\cap\{\mathbf{m}\} = \emptyset$, $|\mathrm{E}\{\tilde{Q}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{m},\tilde{\mathbf{m}},\mathbf{j})\} - \kappa_1^{2a}\prod_{t=1}^4\{\mathrm{E}(x_{1,j_t}^2)\}^a| = 0$ by Condition 3.2.2*; if $\{\mathbf{i}\}\cap\{\mathbf{m}\} \neq \emptyset$, $|\{\mathbf{i}\}\cup\{\mathbf{m}\}| \le a_1 + a_2 - 2 - 1$, thus $|\mathrm{E}(\tilde{T}_{k,a_1,a_2}^2)_{(1)} - \{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2| = O(n^{a_1+a_2-3}p^4)$ by Condition 3.2.1.

Step II: $\mathrm{E}(T_{k,a_1,a_2}^2)_{(2)}$ We note that for $j_1 \neq j_2$, $\mathrm{E}(x_{1,j_1}x_{1,j_2}) = 0$, and for any additional index j_3, we have $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}) = 0$ under Condition 3.2.2*. Thus (B.72) and (B.73) still hold here, and we obtain $\mathrm{E}(T_{k,a_1,a_2}^2)_{(2)} = O(n^{a_1+a_2-2}p^2)$.

In summary,

$$\begin{aligned}|\mathrm{var}(T_{k,a_1,a_2})| &\le |\mathrm{E}(\tilde{T}_{k,a_1,a_2}^2)_{(1)} - \{\mathrm{E}(\tilde{T}_{k,a_1,a_2})\}^2| + |\mathrm{E}(T_{k,a_1,a_2}^2)_{(2)}| \\ &= n^{a_1+a_2-2}p^4 O(n^{-1}+p^{-2}).\end{aligned}$$

It follows that $\mathrm{var}(\sum_{k=1}^{n} \pi_{n,k}^2) = O(n^{-1} + p^{-2})$ by the argument at the beginning of Section B.5.6. Therefore Lemma B.1.5 is proved.

B.5.7 Proof of Lemma B.1.6

By Lemma B.1.4,

$$\sum_{k=1}^{n} \mathrm{E}(D_{n,k}^4) = \sum_{k=1}^{n} \sum_{1 \le r_1, r_2, r_3, r_4 \le m} \prod_{l=1}^{4} t_{r_l} \times \mathrm{E}\Big(\prod_{l=1}^{4} A_{n,k,a_{r_l}} \Big). \tag{B.91}$$

To prove Lemma B.1.6, it suffices to show that for given $1 \le k \le n$ and $1 \le r_1, r_2, r_3, r_4 \le m$, we have $\mathrm{E}(\prod_{l=1}^{4} A_{n,k,a_{r_l}}) = O(n^{-2})$.

Similarly to Sections B.5.2 and B.5.6 above, we first illustrate the proof of Lemma B.1.6, when $x_{i,j}$'s are independent. Then in Section B.5.7.1, we prove Lemma B.1.6 under Condition 3.2.2. Last in Section B.5.7.2, we prove Lemma B.1.6 under Condition 3.2.2*.

Proof illustration In this section, we assume that $x_{i,j}$'s are independent and prove $\mathrm{E}(\prod_{l=1}^{4} A_{n,k,a_l}) = O(n^{-2})$ for given integers a_l, $l = 1, \ldots, 4$. By Lemma B.1.4, when $k < a_l$, $A_{n,k,a_l} = 0$. We next focus on $\max_{1 \le l \le 4} a_l \le k \le n$. By Lemma B.1.4, we have

$$\mathrm{E}\Big(\prod_{l=1}^{4} A_{n,k,a_l} \Big) = \Big\{ \prod_{l=1}^{4} c(n, a_l) \Big\}^{1/2} \sum_{\substack{\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1),\, l=1,\ldots,4; \\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in \mathcal{J}}} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8), \tag{B.92}$$

where $\mathbf{i}^{(l)} = (i_1^{(l)}, \ldots, i_{a_l-1}^{(l)})$, $l = 1, \ldots, 4$ represent the tuples satisfying $1 \le i_1^{(l)} \ne \ldots \ne i_{a_l-1}^{(l)} \le n$; $\mathcal{J} = \{(j_1, j_2) : 1 \le j_1 \ne j_2 \le p)$; $\mathbf{j}_8$ represents the tuple

$(j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8)$; and we define

$$Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = \mathrm{E}\Big(\prod_{r=1}^{8} x_{k,j_r}\Big)\,\mathrm{E}\Big(\prod_{t=1}^{a_l-1}\prod_{l=1}^{4} x_{i_t^{(l)},j_{2l-1}} x_{i_t^{(l)},j_{2l}}\Big).$$

We claim that $\mathrm{E}(\prod_{r=1}^{8} x_{k,j_r}) \neq 0$ only when

$$|\{j_t : t = 1, \ldots, 8\}| \leq 4. \tag{B.93}$$

If $|\{j_t : t = 1, \ldots, 8\}| \geq 5$, it implies that one of the j index in $\{j_t : t = 1, \ldots, 8\}$ only appears once. We assume without loss of generality that j_1 only appears once, i.e., $j_1 \notin \{j_t : t = 2, \ldots, 8\}$. Since $x_{k,j}$'s are independent, $\mathrm{E}(\prod_{r=1}^{8} x_{k,j_r}) = \mathrm{E}(x_{k,j_1})\mathrm{E}(\text{all the remaining terms}) = 0$. Thus (B.93) is proved. Similarly to (B.70) and (B.71), we further know $Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) \neq 0$ only when

$$\Big|\bigcup_{l=1}^{4}\{\mathbf{i}^{(l)}\}\Big| \leq \sum_{l=1}^{4}(a_l - 1)/2. \tag{B.94}$$

In summary, combining (B.93) and (B.94), we have

$$\mathrm{E}\Big(\prod_{l=1}^{4} A_{n,k,a_l}\Big) = O(p^{-4} n^{-\frac{1}{2}\sum_{l=1}^{4} a_l} n^{\frac{1}{2}\sum_{l=1}^{4}(a_l-1)} p^4) = O(n^{-2}).$$

B.5.7.1 Proof under Condition 3.2.2

Section B.5.7 proves Lemma B.1.6 when $x_{i,j}$'s are independent. In this section, we further prove Lemma B.1.6 under Condition 3.2.2. We first illustrate the proof idea intuitively, which is similar to Sections B.5.2 and B.5.6.1. Under Condition 3.2.2, $x_{i,j}$'s may be no longer independent, but the dependence between x_{i,j_1} and x_{i,j_2} degenerates exponentially with their distance $|j_1 - j_2|$. To quantitatively examine $|j_1 - j_2|$, we use the threshold of distance D_0 defined in (B.77). Intuitively, when $|j_1 - j_2| > D_0$, x_{i,j_1} and x_{i,j_2} are "asymptotically independent" with similar properties to those under the independence case in Section B.5.7. The following proof will provide comprehensive discussions based on D_0.

We next present the detailed proof of Lemma B.1.6. Note that to prove Lemma B.1.6, by the analysis at the beginning of Section B.5.7, it suffices to show that $\mathrm{E}(\prod_{l=1}^4 A_{n,k,a_l}) = O(n^{-2})$. Recall that we can write (B.92) and have $\prod_{l=1}^4 c^{1/2}(n, a_l) = \Theta(p^{-4} n^{-\frac{1}{2}\sum_{l=1}^4 a_l})$. It remains to show

$$\sum_{\substack{\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l-1),\, l=1,\ldots,4; \\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in \mathcal{J}}} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = O(p^4 n^{\frac{1}{2}\sum_{l=1}^4 (a_l-1)}). \tag{B.95}$$

To prove (B.95), we show the order of (B.95) in n and p respectively in the following two steps.

Step I: order of n. We show for any fixed $\mathbf{j}_8 = (j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8)$,

$$\left| \sum_{\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l-1),\, l=1,\ldots,4} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) \right| = O(n^{\frac{1}{2}\sum_{l=1}^4 (a_l-1)}). \tag{B.96}$$

We note that $Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) \neq 0$ only if (B.94) holds. Too see this, suppose one index i_1 only appears once in the four sets $\{\mathbf{i}^{(1)}\}, \{\mathbf{i}^{(2)}\}, \{\mathbf{i}^{(3)}\}, \{\mathbf{i}^{(4)}\}$. For example $i_1 \in \{\mathbf{i}^{(1)}\}$, but $i_1 \notin \cup_{l=2}^4 \{\mathbf{i}^{(l)}\}$. Then

$$Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = \mathrm{E}(x_{i_1,j_1} x_{i_1,j_2}) \times \mathrm{E}(\text{the remaining terms}) = 0,$$

Therefore by (B.94) and Condition 3.2.1,

$$(\text{B.96}) = O(n^{\frac{1}{2}\sum_{l=1}^4 (a_l-1)}). \tag{B.97}$$

Step II: order of p. To prove (B.95), it remains to show that for a given tuple $(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)})$,

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = O(p^4). \tag{B.98}$$

Let μ be a positive constant same as in (B.77). Define an event $B_J^c = \{Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = O(p^{-(8+\mu)})\}$ and let B_J represent the complement set of B_J^c correspondingly. Note that

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) \times \mathbf{1}_{B_J^c} = O(p^8 p^{-(8+\mu)}) = o(1).$$

Moreover by Condition 3.2.1, $Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = O(1)$ always holds. Thus to prove (B.98), it remains to show

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in\mathcal{J}} \mathbf{1}_{B_J} = O(p^4). \tag{B.99}$$

We write the ordered version of $\mathbf{j}_8 = (j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8)$ as $\tilde{\mathbf{j}}_8 = (\tilde{j}_1, \tilde{j}_2, \tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6, \tilde{j}_7, \tilde{j}_8)$, where the indexes satisfy $\tilde{j}_1 \le \tilde{j}_2 \le \tilde{j}_3 \le \tilde{j}_4 \le \tilde{j}_5 \le \tilde{j}_6 \le \tilde{j}_7 \le \tilde{j}_8$. To facilitate the proof, we first introduce three claims below, which will be proved later. In particular, for given $\mathbf{j}_8$, if $\mathbf{1}_{B_J} = 1$, the corresponding ordered tuple $\tilde{\mathbf{j}}_8$ of $\mathbf{j}_8$ satisfies the following three claims with D_0 defined in (B.77).

- **Claim 1** : For any index $\tilde{j}_k \in \tilde{\mathbf{j}}_8$, if it has two neighbors $\tilde{j}_{k-1}$ and $\tilde{j}_{k+1}$, its distances with the two neighbors $\tilde{j}_{k-1}$ and $\tilde{j}_{k+1}$ can not be bigger than D_0 together. That is, at least one of $|\tilde{j}_{k-1} - \tilde{j}_k| \le D_0$ and $|\tilde{j}_k - \tilde{j}_{k+1}| \le D_0$ is true. For $\tilde{j}_1$ and $\tilde{j}_8$ with only one neighbor, they satisfy $|\tilde{j}_1 - \tilde{j}_2| \le D_0$ and $|\tilde{j}_7 - \tilde{j}_8| \le D_0$.

- **Claim 2** : For a pair of indexes $(\tilde{j}_{k-1}, \tilde{j}_k)$ in $\tilde{\mathbf{j}}_8$, when $\tilde{j}_{k-1} \ne \tilde{j}_k$, if it has two neighbors $\tilde{j}_{k-2}$ and $\tilde{j}_{k+1}$, the distances of the pair with the two neighbors can not be bigger than D_0 together. That is, at least one of $|\tilde{j}_{k-2} - \tilde{j}_{k-1}| \le D_0$ and $|\tilde{j}_k - \tilde{j}_{k+1}| \le D_0$ holds. For the pairs $(\tilde{j}_1, \tilde{j}_2)$ and $(\tilde{j}_7, \tilde{j}_8)$ with only one neighbor, when $\tilde{j}_1 \ne \tilde{j}_2$ and $\tilde{j}_7 \ne \tilde{j}_8$, they satisfy $|\tilde{j}_2 - \tilde{j}_3| \le D_0$ and $|\tilde{j}_6 - \tilde{j}_7| \le D_0$.

Claim 3 :

(a) For any given $\{\tilde{j}_4, \tilde{j}_5, \tilde{j}_6, \tilde{j}_7, \tilde{j}_8\}$,

$$\sum_{\tilde{j}_1, \tilde{j}_2, \tilde{j}_3} \mathbf{1}_{B_J \cap \{\tilde{j}_1 = \tilde{j}_2\}} = O(p^2), \quad \sum_{\tilde{j}_1, \tilde{j}_2, \tilde{j}_3} \mathbf{1}_{B_J \cap \{\tilde{j}_1 \ne \tilde{j}_2\}} = O(pD_0^2).$$

(b) For any given $\{\tilde{j}_1, \tilde{j}_2, \tilde{j}_3, \tilde{j}_4, \tilde{j}_5\}$,

$$\sum_{\tilde{j}_6, \tilde{j}_7, \tilde{j}_8} \mathbf{1}_{B_J \cap \{\tilde{j}_7 = \tilde{j}_8\}} = O(p^2), \quad \sum_{\tilde{j}_6, \tilde{j}_7, \tilde{j}_8} \mathbf{1}_{B_J \cap \{\tilde{j}_7 \ne \tilde{j}_8\}} = O(pD_0^2).$$

Given three claims above, we show (B.99) by discussing different cases.

1. When both $\tilde{j}_1 \ne \tilde{j}_2$ and $\tilde{j}_7 \ne \tilde{j}_8$, by Claim 3, we know the summation over indexes $(\tilde{j}_1, \tilde{j}_2, \tilde{j}_3)$ is of order pD_0^2 and the summation over indexes $(\tilde{j}_6, \tilde{j}_7, \tilde{j}_8)$ is also of order pD_0^2. Then we consider $(\tilde{j}_4, \tilde{j}_5)$. When $|\tilde{j}_4 - \tilde{j}_5| \le D_0$, the summation is of order $(pD_0^2) \times pD_0 \times pD_0^2 = p^3 D_0^5 = p^4$. When $|\tilde{j}_4 - \tilde{j}_5| > D_0$, applying Claim 1 on $\tilde{j}_4$ and $\tilde{j}_5$ respectively, we know $|\tilde{j}_3 - \tilde{j}_4| \le D_0$ and $|\tilde{j}_5 - \tilde{j}_6| \le D_0$ hold. Therefore, the summation is of order $pD_0^2 \times D_0 \times p \times D_0 \times pD_0^2 = p^3 D_0^6 = p^4$. In summary,

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in \mathcal{J}} \mathbf{1}_{B_J \cap \{\tilde{j}_1 \ne \tilde{j}_2, \tilde{j}_7 \ne \tilde{j}_8\}} = O(p^4).$$

2. When only one of $\tilde{j}_1 \ne \tilde{j}_2$ and $\tilde{j}_7 \ne \tilde{j}_8$ holds, without loss of generality, we consider $\tilde{j}_1 = \tilde{j}_2$ and $\tilde{j}_7 \ne \tilde{j}_8$.

   (a) When $|\tilde{j}_2 - \tilde{j}_3| > D_0$, applying Claim 1 on $\tilde{j}_3$, we know $|\tilde{j}_3 - \tilde{j}_4| \le D_0$. Then consider the pair $(\tilde{j}_3, \tilde{j}_4)$. If $\tilde{j}_3 = \tilde{j}_4$, by Claim 1, $|\tilde{j}_5 - \tilde{j}_4| \leq D_0$ or $|\tilde{j}_5 - \tilde{j}_6| \leq D_0$ holds. As $\tilde{j}_7 \neq \tilde{j}_8$, by Claim 3, the summation over $(\tilde{j}_6, \tilde{j}_7, \tilde{j}_8)$ is of order pD_0^2. Therefore, the total summation order is $O(p \times p \times D_0 \times pD_0^2) = O(p^4)$. If $\tilde{j}_3 \neq \tilde{j}_4$, applying Claim 2 on the pair $(\tilde{j}_3, \tilde{j}_4)$, we know $|\tilde{j}_4 - \tilde{j}_5| \leq D_0$ as we discuss $|\tilde{j}_2 - \tilde{j}_3| > D_0$. Also, as $\tilde{j}_7 \neq \tilde{j}_8$, by Claim 3, the summation order over $(\tilde{j}_6, \tilde{j}_7, \tilde{j}_8)$ is $O(pD_0^2)$. Thus the total order of summation is $O(pD_0 pD_0^2 pD_0^2) = O(p^4)$. In summary,

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbf{1}_{\left\{B_J \cap \{\text{one of } \tilde{j}_1 \neq \tilde{j}_2 \text{ or } \tilde{j}_7 \neq \tilde{j}_8,\, |\tilde{j}_2 - \tilde{j}_3| > D_0\}\right\}} = O(p^4).$$

(b) When $|\tilde{j}_2 - \tilde{j}_3| \leq D_0$, the summation over $\tilde{j}_1, \tilde{j}_2, \tilde{j}_3$ is of order pD_0. Then we consider $\tilde{j}_4, \tilde{j}_5$. If $|\tilde{j}_4 - \tilde{j}_5| \leq D_0$, the summation over $\tilde{j}_1, \tilde{j}_2, \tilde{j}_3, \tilde{j}_4, \tilde{j}_5$ is of order $pD_0 pD_0 = p^2 D_0^2$. As $\tilde{j}_7 \neq \tilde{j}_8$, by Claim 3, we know the summation order of $\tilde{j}_6, \tilde{j}_7, \tilde{j}_8$ is pD_0^2. Then the total summation order of this case is $O(1) p^2 D_0^2 p D_0^2 = O(p^4)$. If $|\tilde{j}_4 - \tilde{j}_5| > D_0$, applying Claim 1 on $\tilde{j}_4$ and $\tilde{j}_5$ respectively, we have $|\tilde{j}_3 - \tilde{j}_4| \leq D_0$ and $|\tilde{j}_5 - \tilde{j}_6| \leq D_0$. Also, as $\tilde{j}_7 \neq \tilde{j}_8$, by Claim 3, we know the summation order of $\tilde{j}_6, \tilde{j}_7, \tilde{j}_8$ is $O(pD_0^2)$. Then the total summation order is $O(1) pD_0 \times D_0 p D_0 \times pD_0^2 = O(p^4)$. In summary,

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbf{1}_{\left\{B_J \cap \{\text{one of } \tilde{j}_1 \neq \tilde{j}_2 \text{ or } \tilde{j}_7 \neq \tilde{j}_8, |\tilde{j}_2 - \tilde{j}_3| \leq D_0\}\right\}} = O(p^4).$$

3. When both $\tilde{j}_1 = \tilde{j}_2$ and $\tilde{j}_7 = \tilde{j}_8$, then we consider $(\tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6)$.

(a) If the number of distinct elements in $\{\tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6\}$ is smaller and equal to 2, the order of summation over $\tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6$ is $O(p^2)$. We use $|\{\tilde{j}_3, \tilde{j}_4, \tilde{j}_5, \tilde{j}_6\}| \leq$

2 to represent this case, then

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbf{1}_{\left\{B_J\cap\{\tilde{j}_1=\tilde{j}_2,\,\tilde{j}_7=\tilde{j}_8,\,|\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}|\le 2\}\right\}} = O(p^4).$$

(b) When the number of distinct elements in $\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}$ is 3, we write $|\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}|=3$ to represent this case. Then two of $\tilde{j}_3\neq\tilde{j}_4$, $\tilde{j}_4\neq\tilde{j}_5$ and $\tilde{j}_5\neq\tilde{j}_6$ hold. We consider without loss of generality $\tilde{j}_3\neq\tilde{j}_4$, $\tilde{j}_4\neq\tilde{j}_5$ and $\tilde{j}_5=\tilde{j}_6$. We apply Claim 2 on the pair $(\tilde{j}_3,\tilde{j}_4)$ and Claim 1 on $\tilde{j}_3$. Then at least two of $|\tilde{j}_2-\tilde{j}_3|\le D_0$, $|\tilde{j}_3-\tilde{j}_4|\le D_0$ and $|\tilde{j}_4-\tilde{j}_5|\le D_0$ holds. Thus the summation order is $O(pD_0^2p^2)=O(p^4)$. In summary,

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbf{1}_{\left\{B_J\cap\{\tilde{j}_1=\tilde{j}_2,\,\tilde{j}_7=\tilde{j}_8,\,|\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}|=3\}\right\}} = O(p^4).$$

(c) When the number of distinct elements in $\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}$ is 4, we write $|\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}|=4$ to represent this case, and we know $\tilde{j}_3\neq\tilde{j}_4$, $\tilde{j}_4\neq\tilde{j}_5$ and $\tilde{j}_5\neq\tilde{j}_6$. Applying Claim 2 on the pair $(\tilde{j}_3,\tilde{j}_4)$, and applying Claim 1 on the two single indexes $\tilde{j}_3$ and $\tilde{j}_4$ respectively, we know at least two of $|\tilde{j}_2-\tilde{j}_3|\le D_0$, $|\tilde{j}_3-\tilde{j}_4|\le D_0$ and $|\tilde{j}_4-\tilde{j}_5|\le D_0$ hold. Therefore the summation over $(\tilde{j}_1,\tilde{j}_2,\tilde{j}_3,\tilde{j}_4,\tilde{j}_5)$ is of order $O(p\times pD_0^2)=O(p^2D_0^2)$. Then applying Claim 1 on $\tilde{j}_6$, we know at least one of $|\tilde{j}_5-\tilde{j}_6|\le D_0$ and $|\tilde{j}_6-\tilde{j}_7|\le D_0$ holds. Then the total order of summation for this part is $O(p^2D_0^2\times pD_0)=O(p^4)$, that is,

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in\mathcal{J}}} \mathbf{1}_{B_J\cap\{\tilde{j}_1=\tilde{j}_2,\,\tilde{j}_7=\tilde{j}_8,\,|\{\tilde{j}_3,\tilde{j}_4,\tilde{j}_5,\tilde{j}_6\}|=4\}} = O(p^4).$$

Combining the results obtained, we know (B.99) is proved. Thus to prove (B.98), it remains to prove the three claims above.

By the definition of $Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8)$ in Section B.5.7,

$$\Big|Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8)\Big| \leq C\Big|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\Big|.$$

Then it is sufficient to show that for given $\mathbf{j}_8$, when the ordered version $\tilde{\mathbf{j}}_8$ of $\mathbf{j}_8$ does not follow the three claims,

$$\Big|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\Big| = O(p^{-(8+\mu)}). \tag{B.100}$$

Proof of Claim 1:

(1) When the index $\tilde{j}_k$ has two neighbors, we give the proof by an example of $k = 3$. All the other cases can be obtained following similar analysis without loss of generality. Suppose $\tilde{j}_3$'s distances between its neighbors $\tilde{j}_2$ and $\tilde{j}_4$ are both bigger than D_0, i.e., $|\tilde{j}_2 - \tilde{j}_3| > D_0$ and $|\tilde{j}_3 - \tilde{j}_4| > D_0$. Then by Conditions 3.2.1, 3.2.2, and the α-mixing inequality in Lemma B.5.1,

$$\begin{aligned}
& \Big|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\Big| \\
= \;& \Big|\mathrm{cov}\Big(\prod_{t=1}^{3} x_{k,\tilde{j}_t}\,,\, \prod_{t=4}^{8} x_{k,\tilde{j}_t}\Big) + \mathrm{E}\Big(\prod_{t=1}^{3} x_{k,\tilde{j}_t}\Big) \times \mathrm{E}\Big(\prod_{t=4}^{8} x_{k,\tilde{j}_t}\Big)\Big| \\
\leq \;& C\delta^{\frac{D_0\epsilon}{2+\epsilon}} + C \times |\mathrm{cov}(x_{k,\tilde{j}_1}x_{k,\tilde{j}_2}\;,\; x_{k,\tilde{j}_3}) + \mathrm{E}(x_{k,\tilde{j}_1}x_{k,\tilde{j}_2})\mathrm{E}(x_{k,\tilde{j}_3})| \\
= \;& O(p^{-(8+\mu)}) + C \times |\mathrm{cov}(x_{k,\tilde{j}_1}x_{k,\tilde{j}_2}\;,\; x_{k,\tilde{j}_3})| \\
= \;& O(p^{-(8+\mu)}).
\end{aligned}$$

Thus (B.100) holds.

(2) For $\tilde{j}_1$ and $\tilde{j}_8$ with only one neighbor, we give the proof on $\tilde{j}_1$, while $\tilde{j}_8$ can be

proved similarly. By Conditions 3.2.1, 3.2.2, and Lemma B.5.1,

$$\begin{aligned}\Big|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\Big| &= \Big|\mathrm{cov}\Big(x_{k,\tilde{j}_1}\,,\,\prod_{t=2}^{8} x_{k,\tilde{j}_t}\Big) + \mathrm{E}(x_{k,\tilde{j}_1}) \times \mathrm{E}\Big(\prod_{t=2}^{8} x_{k,\tilde{j}_t}\Big)\Big| \\ &\le C\delta^{\frac{D_0\epsilon}{2+\epsilon}} + 0 \qquad (\ \mathrm{E}(x_{k,\tilde{j}_1}) = 0\ ) \\ &= O(p^{-(8+\mu)}).\end{aligned}$$

Thus (B.100) also holds.

Proof of Claim 2:

(1) When the pair $(\tilde{j}_{k-1}, \tilde{j}_k)$ has two neighbors, we give the proof by the example when $k = 5$, i.e., we consider the pair $(\tilde{j}_4, \tilde{j}_5)$. The other cases can be proved similarly without loss of generality. Suppose $\tilde{j}_4 \neq \tilde{j}_5$ with $|\tilde{j}_3 - \tilde{j}_4| > D_0$ and $|\tilde{j}_5 - \tilde{j}_6| > D_0$. As $\mathrm{E}(x_{k,\tilde{j}_4} x_{k,\tilde{j}_5}) = 0$ under H_0, by Conditions 3.2.1 and 3.2.2, and Lemma B.5.1, we have

$$\begin{aligned}&\Big|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\Big| \\ &= \Big|\mathrm{cov}\Big(\prod_{t=1}^{3} x_{k,\tilde{j}_t}\,,\,\prod_{t=4}^{8} x_{k,\tilde{j}_t}\Big) + \mathrm{E}\Big(\prod_{t=1}^{3} x_{k,\tilde{j}_t}\Big) \times \mathrm{E}\Big(\prod_{t=4}^{8} x_{k,\tilde{j}_t}\Big)\Big| \\ &\le C\delta^{\frac{D_0\epsilon}{2+\epsilon}} + \Big|\mathrm{E}\Big(\prod_{t=1}^{3} x_{k,\tilde{j}_t}\Big) \times \Big\{\mathrm{cov}\Big(\prod_{t=4}^{5} x_{k,\tilde{j}_t}\ ,\ \prod_{t=6}^{8} x_{k,\tilde{j}_t}\Big) + \mathrm{E}\Big(\prod_{t=4}^{5} x_{\tilde{j}_t}\Big)\mathrm{E}\Big(\prod_{t=6}^{8} x_{k,\tilde{j}_t}\Big)\Big\}\Big| \\ &= C\delta^{\frac{D_0\epsilon}{2+\epsilon}} + \Big|\mathrm{E}(x_{k,\tilde{j}_1} x_{k,\tilde{j}_2} x_{\tilde{j}_3}) \times \{\mathrm{cov}(x_{k,\tilde{j}_4} x_{k,\tilde{j}_5}\ ,\ x_{k,\tilde{j}_6} x_{k,\tilde{j}_7} x_{k,\tilde{j}_8}) + 0\}\Big| \\ &\le C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(p^{-(8+\mu)}).\end{aligned}$$

Thus (B.100) holds.

(2) For the pairs $(\tilde{j}_1, \tilde{j}_2)$ and $(\tilde{j}_7, \tilde{j}_8)$ with only one neighbor, we give the proof on $(\tilde{j}_1, \tilde{j}_2)$, while the proof on $(\tilde{j}_7, \tilde{j}_8)$ can be obtained similarly. If $\tilde{j}_1 \neq \tilde{j}_2$ and $|\tilde{j}_2 - \tilde{j}_3| > D_0$, as $\mathrm{E}(x_{k,\tilde{j}_1} x_{k,\tilde{j}_2}) = 0$ under H_0, by Conditions 3.2.1 and 3.2.2, and the

α-mixing inequality in Lemma B.5.1, we have

$$\begin{aligned}\Big|\mathrm{E}\Big(\prod_{t=1}^{8} x_{k,\tilde{j}_t}\Big)\Big| &= \Big|\mathrm{cov}\Big(\prod_{t=1}^{2} x_{k,\tilde{j}_t}\,,\ \prod_{t=3}^{8} x_{\tilde{j}_t}\Big) + \mathrm{E}\Big(\prod_{t=1}^{2} x_{k,\tilde{j}_t}\Big)\mathrm{E}\Big(\prod_{t=3}^{8} x_{\tilde{j}_t}\Big)\Big| \\ &\le C\delta^{\frac{D_0\epsilon}{2+\epsilon}} = O(Cp^{-(8+\mu)}).\end{aligned}$$

Thus (B.100) holds.

Proof of Claim 3: The Claim 3 (a) is obtained by applying Claim 1 on the $\tilde{j}_1$ and Claim 2 on the pair $(\tilde{j}_1, \tilde{j}_2)$ when $\tilde{j}_1 \neq \tilde{j}_2$. The Claim 3 (b) is also obtained similarly.

B.5.7.2 Proof under Condition 3.2.2*

In this section, we prove Lemma B.1.6 by substituting Condition 3.2.2 with Condition 3.2.2*. Similarly to Section B.5.6.2, the proof under Condition 3.2.2* follows similarly to the proof under the independence case in Section B.5.7. In particular, we note that Condition 3.2.2* implies that if one of the indexes in $\{j_1, \ldots, j_8\}$ only appears once, $\mathrm{E}(\prod_{r=1}^{8} x_{k,j_r}) = 0$. Therefore when $\mathrm{E}(\prod_{r=1}^{8} x_{k,j_r}) \neq 0$, (B.93) holds. Also following similar analysis, we know (B.94) holds by Condition 3.2.2* and $\mathrm{E}(x_{1,j_1}x_{1,j_2}) = 0$ for $j_1 \neq j_2$. Combining (B.93) and (B.94), Lemma B.1.6 is proved.

B.5.8 Proof of Lemma B.1.7

For easy illustration, we first prove Lemma B.1.7 when $m = 1$ and next present the proof for $m > 1$.

Proof for $m = 1$ Specifically, in this section, we prove

$$\Big|P\Big(\frac{\hat{M}_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \le 2z\Big) - P\Big(\frac{\hat{M}_n}{n} > y_p\Big)P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \le 2z\Big)\Big| \to 0.$$

Note that by definitions in (B.6) and (B.7),

$$
\begin{aligned}
& P\Big(\frac{\hat{M}_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \leq 2z\Big) && \text{(B.101)} \\
= \ & P\Big(\max_{1\leq l\leq q} (\hat{G}_l)^2 > ny_p, (\sigma(a)P_a^n)^{-1}\sum_{m=1}^{q} U_m^a \leq z\Big) \\
= \ & P\Big(\Big\{ \cup_{l=1}^q \{(\hat{G}_l)^2 > ny_p\}\Big\} \cap \Big\{(\sigma(a)P_a^n)^{-1}\sum_{m=1}^{q} U_m^a \leq z\Big\}\Big).
\end{aligned}
$$

Define the events $E_l = \{(\hat{G}_l)^2 > ny_p\} \cap \{(\sigma(a)P_a^n)^{-1}\sum_{m=1}^q U_m^a \leq z\}$, and then we have

$$
\text{(B.101)} = P(\cup_{l=1}^q E_l). \tag{B.102}
$$

We next examine the upper and lower bounds of (B.102). Particularly, using the Bonferroni's inequality, for any even number $d < [q/2]$, we obtain

$$
\begin{aligned}
\sum_{s=1}^{d}(-1)^{s-1} \sum_{1\leq l_1<\ldots<l_s\leq q} P(\cap_{t=1}^s E_{l_t}) &\leq P(\ \cup_{l=1}^q E_l\ ) && \text{(B.103)} \\
&\leq \sum_{s=1}^{d-1}(-1)^{s-1} \sum_{1\leq l_1<\ldots<l_s\leq q} P(\cap_{t=1}^s E_{l_t}).
\end{aligned}
$$

We consider $d = O(\log^{1/5} p)$ below. The following proof proceeds by examining the upper and lower bounds of $P(\cap_{t=1}^s E_{l_t})$ first and combining them based on (B.103).

To facilitate the discussion, we define some notation. Let

$$
H_d = \sum_{s=1}^{d}(-1)^{s-1} \sum_{1\leq l_1<\ldots<l_s\leq q} P(\cap_{t=1}^s\{(\hat{G}_{l_t})^2 > ny_p\}).
$$

By the Bonferroni's inequality, we have

$$
H_d \leq P(\ \cup_{l=1}^q\{(\hat{G}_l)^2 > ny_p\}\ ) \leq H_{d-1}. \tag{B.104}
$$

Given $l_1, \ldots, l_s$, we define two index sets: $I_s = \{(j_{l_t}^1, j_{l_t}^2), 1 \le t \le s\}$ and correspondingly

$$L_{I_s} = \{(j_1, j_2) : \ (j_1, j_2) \cap (u, t) \neq \emptyset, (u, t) \in I_s \text{ and } (j_1, j_2) \in L\}, \tag{B.105}$$

where L is defined in (B.5). (B.105) suggests that L_{I_s} contains all the index pairs that have overlap with the index pairs in I_s. Note that the definitions of I_s and L_{I_s} depend on the given indexes $l_1, \ldots, l_s$; for the simplicity of notation, we write I_s and L_{I_s} in this proof without ambiguity. It follows that

$$\sum_{m=1}^{q} U_m^a = \sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^a + \sum_{(j_l^1, j_l^2) \in L \backslash L_{I_s}} U_l^a. \tag{B.106}$$

The cardinality of L_{I_s} is no greater than $2ps$ by construction. Furthermore, $2ps \le 2pd$ as $s \le d$. Note that the indexes in I_s and $L \backslash L_{I_s}$ have no intersection. By this construction and the independence assumption in Condition 3.2.3, for any finite integers $a_1, a_2 \ge 1$, we know

$$\{U_l^{a_1}, \ (j_l^1, j_l^2) \in I_s\} \quad \text{and} \quad \{U_l^{a_2}, \ (j_l^1, j_l^2) \in L \backslash L_{I_s}\}$$

are independent.

We next examine the upper bound of $P(\cap_{t=1}^s E_{l_t})$. By the definition of E_l and

(B.106),

$$
\begin{aligned}
&P(\cap_{t=1}^{s} E_{l_t}) \qquad\qquad \text{(B.107)}\\
&= P\Big(\bigcap_{t=1}^{s}\Big\{\Big\{(\sigma(a)P_a^n)^{-1}\sum_{m=1}^{q} U_m^a \le z\Big\}\bigcap\{(\hat{G}_{l_t})^2 > ny_p\}\Big\}\Big)\\
&= P\Big(\bigcap_{t=1}^{s}\Big\{\Big\{(\sigma(a)P_a^n)^{-1}\Big[\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a + \sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}} U_l^a\Big] \le z\Big\}\\
&\qquad \bigcap\{(\hat{G}_{l_t})^2 > ny_p\}\Big\}\Big).
\end{aligned}
$$

Let Γ_p represent a number of order $\Theta\{(\log p)^{-1/2}\}$ and we have

$$
\begin{aligned}
&\Big\{(\sigma(a)P_a^n)^{-1}\Big(\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a + \sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}} U_l^a\Big) \le z\Big\}\\
&\subseteq \Big\{(\sigma(a)P_a^n)^{-1}\Big|\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a\Big| \ge \Gamma_p\Big\}\bigcup\Big\{(\sigma(a)P_a^n)^{-1}\sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}} U_l^a \le \Gamma_p + z\Big\}.
\end{aligned}
$$

Thus (B.107) has the following upper bound,

$$
\begin{aligned}
\text{(B.107)} \le\; & P\Big(\Big\{\cap_{t=1}^{s}\{(\hat{G}_{l_t})^2 > ny_p\}\Big\}\bigcap\Big\{(\sigma(a)P_a^n)^{-1}\Big|\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a\Big| \ge \Gamma_p\Big\}\Big)\\
&+ P\Big(\Big\{\cap_{t=1}^{s}\{(\hat{G}_{l_t})^2 > ny_p\}\Big\}\bigcap\Big\{(\sigma(a)P_a^n)^{-1}\sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}} U_l^a \le \Gamma_p + z\Big\}\Big).
\end{aligned}
$$

In addition, we note that $\{\hat{G}_l,\ (j_l^1, j_l^2) \in I_s\}$ and $\{U_l^a,\ (j_l^1, j_l^2) \in L\backslash L_{I_s}\}$ are independent, because of $I_s \cap (L\backslash L_{I_s}) = \emptyset$ by the construction and the independence assumption in Condition 3.2.3. It follows that

$$
\text{(B.107)} \le P_s + P_{ys}P_{+z}, \tag{B.108}
$$

where for simplicity we define

$$
\begin{aligned}
P_s &= P\Big(\Big\{(\sigma(a)P_a^n)^{-1}\Big|\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a\Big| \geq \Gamma_p\Big\}\Big), &\text{(B.109)}\\
P_{ys} &= P\Big(\bigcap_{t=1}^{s}\{(\hat{G}_{l_t})^2 > ny_p\}\Big),\\
P_{+z} &= P\Big(\Big\{(\sigma(a)P_a^n)^{-1}\sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}} U_l^a \leq \Gamma_p + z\Big\}\Big).
\end{aligned}
$$

Note that although the notation P_{ys}, P_{+z} and P_s in (B.109) suppress their dependence on the specific choice of $(l_1,\ldots,l_s)$, this will not influence the proof due to the i.i.d. assumption in Condition 3.2.3.

Similarly we examine the lower bound of $P(\cap_{t=1}^s E_{l_t})$. In particular,

$$
\begin{aligned}
&\Big\{(\sigma(a)P_a^n)^{-1}\sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}} U_l^a \leq z - \Gamma_p\Big\}\\
\subseteq\ &\Big\{(\sigma(a)P_a^n)^{-1}\Big|\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a\Big| \geq \Gamma_p\Big\}\bigcup\Big\{(\sigma(a)P_a^n)^{-1}\sum_{m=1}^{q} U_m^a \leq z\Big\}.
\end{aligned}
$$

Then (B.107) has the following lower bound,

$$
\begin{aligned}
\text{(B.107)} \geq\ & -P\Big(\Big\{\cap_{t=1}^s\{(\hat{G}_{l_t})^2 > ny_p\}\Big\}\bigcap\Big\{(\sigma(a)P_a^n)^{-1}\Big|\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a\Big| \geq \Gamma_p\Big\}\Big)\\
&+P\Big(\Big\{\cap_{t=1}^s\{(\hat{G}_{l_t})^2 > ny_p\}\Big\}\bigcap\Big\{(\sigma(a)P_a^n)^{-1}\sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}} U_l^a \leq z-\Gamma_p\Big\}\Big).
\end{aligned}
$$

Similarly to (B.108), as $\{\hat{G}_l,\ (j_l^1,j_l^2)\in I_s\}$ and $\{U_l^a,\ (j_l^1,j_l^2)\in L\backslash L_{I_s}\}$ are independent, we obtain

$$
\text{(B.107)} \geq P_{ys}\times P_{-z} - P_s, \tag{B.110}
$$

where P_{ys} and P_s are defined same as in (B.109), and we define

$$P_{-z} = P\Big((\sigma(a)P_a^n)^{-1} \sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^a \le z - \Gamma_p\Big).$$

We have obtained the upper and lower bounds of $P(\cap_{t=1}^s E_{l_t})$ in (B.108) and (B.110) respectively. We next prove that P_{+z} in (B.108) and P_{-z} in (B.110) are close in the sense that there exists some constant $C > 0$,

$$|P_{+z} - P_z| \le C \times \Gamma_p \quad \text{and} \quad |P_{-z} - P_z| \le C \times \Gamma_p, \tag{B.111}$$

where we define $P_z = P((\sigma(a)P_a^n)^{-1} \sum_{m=1}^q U_m^a \le z)$. To obtain (B.111), we note that $\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a$ is a summation over index pairs in L_{I_s}, and L_{I_s} is of size $2ps$, which is $o(p^2)$ as $s \le d$ and $d = O(\log^5 p)$. Following similar analysis of $\tilde{\mathcal{U}}^*(a)/\sigma(a) \xrightarrow{P} 0$ in Lemma B.1.1, we know $(\sigma(a)P_a^n)^{-1}\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^a \xrightarrow{P} 0$. Moreover, by $\tilde{\mathcal{U}}(a) = 2(P_a^n)^{-1}\sum_{l=1}^q U_l^a$ in (B.6), $\Gamma_p = \Theta(\log^{-1/2} p)$ and the convergence result in (B.4), we have for given z,

$$|P_{+z} - \Phi(2z + 2\Gamma_p)| \le C\Gamma_p,\ |P_{-z} - \Phi(2z - 2\Gamma_p)| \le C\Gamma_p,\ |P_z - \Phi(2z)| \le C\Gamma_p.$$

As $|\Phi(2z + 2\Gamma_p) - \Phi(2z)| \le C\Gamma_p$ for given z, $|P_{+z} - P_z| \le |P_{+z} - \Phi(2z + 2\Gamma_p)| + |\Phi(2z+2\Gamma_p) - \Phi(2z)| + |P_z - \Phi(2z)| \le C\Gamma_p$. Similarly, as $|\Phi(2z-2\Gamma_p) - \Phi(2z)| \le C\Gamma_p$, $|P_{-z} - P_z| \le C\Gamma_p$. Therefore (B.111) is obtained.

In summary, given (B.108), (B.110) and (B.111), we have

$$|P(\cap_{t=1}^s E_{l_t}) - P_{ys} \times P_z| \le P_s + C \times \Gamma_p \times P_{ys}.$$

Given the above property of $P(\cap_{t=1}^s E_{l_t})$, we next derive an upper bound of (B.102) based on the relationship in (B.103). Specifically,

$$
\begin{aligned}
& P(\cup_{l=1}^{q} E_l) \\
\le & \sum_{s=1}^{d-1} (-1)^{s-1} \sum_{1 \le l_1 < \ldots < l_s \le q} P(\cap_{t=1}^{s} E_{l_t}) \\
\le & \sum_{s=1}^{d-1} (-1)^{s-1} \sum_{1 \le l_1 < \ldots < l_s \le q} \{P_{ys} P_z + (-1)^{s-1} \times [C\Gamma_p \times P_{ys} + P_s]\} \\
\le & H_{d-1} \times P_z + \sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} (C \times \Gamma_p \times P_{ys} + P_s),
\end{aligned}
\tag{B.112}
$$

where the last inequality uses the notation in (B.104), that is,

$$
H_{d-1} = \sum_{s=1}^{d-1} (-1)^{s-1} \sum_{1 \le l_1 < \ldots < l_s \le q} P_{ys}, \tag{B.113}
$$

and the fact that P_z does not depend on $l_1, \ldots, l_s$ in summation. From (B.104), we know $H_{d-1} \le P_y + |H_{d-1} - H_d|$, where we define

$$
P_y = P\Big(\bigcup_{l=1}^{q} \{(\hat{G}_l)^2 > n y_p\} \Big). \tag{B.114}
$$

As a result, we have

$$
\text{(B.112)} \le P_y \times P_z + |H_{d-1} - H_d| \times P_z + \sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} (C\Gamma_p P_{ys} + P_s).
$$

Next we can obtain $|H_{d-1} - H_d| \times P_z \to 0$, $\sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} \Gamma_p \times P_{ys} \to 0$ and $\sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} P_s \to 0$ by the following three Lemmas B.5.5–B.5.7, respectively.

Lemma B.5.5. *Under the conditions of Theorem 3.2.3, when* $s = O(\log^{1/5} p)$*,*

$$\begin{aligned} & \sum_{1 \le l_1 < \ldots < l_s \le q} P\Big(\bigcap_{t=1}^{s} \{(\hat{G}_{l_t})^2/n \ge 4\log p - \log\log p + y\}\Big) \\ = \; & \frac{1}{s!}\Big(\frac{1}{2\sqrt{2\pi}} e^{-\frac{y}{2}}\Big)^s (1 + o(1)) + o(1). \end{aligned}$$

Proof. *See Section B.2.2 in He et al. (2021e).* ◻

Lemma B.5.6. *Under the conditions of Theorem 3.2.3, when* $d = O(\log^{1/5} p)$*,*

$$\begin{aligned} & \sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} P\Big(\bigcap_{t=1}^{s} \{(\hat{G}_{l_t})^2/n \ge 4\log p - \log\log p + y\}\Big) \\ = \; & \sum_{s=1}^{d-1} \frac{1}{s!}\Big(\frac{1}{2\sqrt{2\pi}} e^{-y/2}\Big)^s \{1 + o(1)\} + o(1). \end{aligned}$$

Proof. *See Section B.2.3 in He et al. (2021e).* ◻

Lemma B.5.7. *Under the conditions of Theorem 3.2.3,*

$$\sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} P\bigg(\Big\{(\sigma(a)P_a^n)^{-1}\Big|\sum_{(j_l^1, j_l^2) \in L_{I_s}} U_l^a\Big| \ge \Gamma_p\Big\}\bigg) \to 0,$$

where L_{I_s} *is defined in* (B.105)*,* $d = O(\log^{1/5} p)$*,* $q = \binom{p}{2}$ *and* $\Gamma_p = \Theta(\log^{-1/2} p)$*.*

Proof. *See Section B.2.4 in He et al. (2021e).* ◻

First, we show $|H_{d-1} - H_d| \times P_z \to 0$. By Lemma B.5.5, when $d \to \infty$,

$$\begin{aligned} |H_{d-1} - H_d| \; &= \; \sum_{1 \le l_1 < \ldots < l_d \le q} P\Big(\bigcap_{t=1}^{d} \{(\hat{G}_{l_t})^2 > n y_p\}\Big) \\ &\le \; C\frac{1}{d!}\Big(\frac{1}{2\sqrt{2\pi}} e^{-y/2}\Big)^d \le Ce \times \Big(\frac{e^{1-y/2}}{2\sqrt{2\pi}d}\Big)^d \to 0, \end{aligned}$$

where the last inequality follows from the fact that $d! \ge e(d/e)^d$. Second, we show that $\sum_{s=1}^{d-1} \sum_{1 \le l_1 < \ldots < l_s \le q} \Gamma_p P_{ys} \to 0$. By the definition of P_{ys} in (B.109), and Lemma

B.5.5, $\sum_{s=1}^{d-1}\sum_{1\le l_1<\ldots<l_s\le q}\Gamma_p P_{ys} = \Gamma_p \sum_{s=1}^{d-1}\frac{1}{s!}(\frac{1}{2\sqrt{2\pi}}e^{-y/2})^s + o(1) \to 0$, where we use $\Gamma_p = \Theta(\log^{-1/2} p) \to 0$ and $\sum_{s=1}^{d-1}\frac{1}{s!}(\frac{1}{2\sqrt{2\pi}}e^{-y/2})^s < \infty$ from $s! \ge e(s/e)^s$. Third, we obtain $\sum_{s=1}^{d-1}\sum_{1\le l_1<\ldots<l_s\le q} P_s \to 0$ directly from Lemma B.5.7 following the notation P_s in (B.109).

In summary, the analysis above shows that $P(\cup_{l=1}^q E_l) \le P_y \times P_z + o(1)$. On the other hand, following similar arguments, we can obtain $P(\cup_{l=1}^q E_l) \ge P_y \times P_z + o(1)$. Therefore, $|P(\cup_{l=1}^q E_l) - P_y \times P_z| \to 0$ is obtained, that is,

$$\Big|P(\cup_{l=1}^q E_l) - P(\ \cup_{l=1}^q\{(\hat{G}_l)^2 > ny_p\}\ )P\Big(\Big\{(\sigma(a)P_a^n)^{-1}\sum_{m=1}^q U_m^a \le z\Big\}\Big)\Big| \to 0.$$

Recall the notation in (B.101) and (B.102). We then know Lemma B.1.7 is proved for $m = 1$.

Proof for $m > 1$ We still use the notation defined in Section B.1.2, where $U_l^{a_r}$ and $\tilde{\mathcal{U}}(a_r)$ for $r = 1, \ldots, m$ follow the definitions in (B.6) and (3.5) respectively. To prove Lemma B.1.7 for $m > 1$, we note that similarly to (B.102), we can write

$$P\Big(\frac{\hat{M}_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a_1)}{\sigma(a_1)} \le 2z_1, \ldots, \frac{\tilde{\mathcal{U}}(a_m)}{\sigma(a_m)} \le 2z_m\Big) = P(\cup_{l=1}^q E_l), \quad \text{(B.115)}$$

where we redefine the events

$$E_l = \bigcap_{r=1}^m \Big\{(\sigma(a_r)P_{a_r}^n)^{-1}\sum_{v=1}^q U_v^{a_r} \le z_r\Big\} \cap \{(\hat{G}_l)^2 > ny_p\}.$$

It follows that (B.103) and (B.104) still hold. For given $l_1, \ldots, l_s$, we define I_s and L_{I_s} same as in (B.105). Then for $r = 1, \ldots, m$, we write

$$\sum_{v=1}^q U_v^{a_r} = \sum_{(j_l^1, j_l^2)\in L_{I_s}} U_l^{a_r} + \sum_{(j_l^1, j_l^2)\in L\setminus L_{I_s}} U_l^{a_r}.$$

By the construction of L_{I_s} and the independence assumption in Condition 3.2.3, we know

$$\cup_{r=1}^{m}\{U_l^{a_r},\ (j_l^1,j_l^2)\in I_s\}\quad\text{and}\quad \cup_{r=1}^{m}\{U_l^{a_r},\ (j_l^1,j_l^2)\in L\backslash L_{I_s}\}$$

are independent.

Similarly to (B.107), given $l_1,\ldots,l_s$, we have

$$\begin{aligned}
&P(\cap_{t=1}^{s}E_{l_t}) \qquad\qquad (B.116)\\
&= P\biggl(\bigcap_{r=1}^{m}\Bigl\{(\sigma(a_r)P_{a_r}^n)^{-1}\Bigl[\sum_{(j_l^1,j_l^2)\in L_{I_s}}U_l^{a_r}+\sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}}U_l^{a_r}\Bigr]\le z_r\Bigr\}\\
&\qquad\cap\{\cap_{t=1}^{s}\{(\hat{G}_{l_t})^2>ny_p\}\}\biggr).
\end{aligned}$$

We take Γ_p same as in Section B.5.8 with $\Gamma_p=\Theta\{(\log p)^{-1/2}\}$. Then for each $r=1,\ldots,m$, we have

$$\begin{aligned}
&\Bigl\{(\sigma(a_r)P_{a_r}^n)^{-1}\Bigl[\sum_{(j_l^1,j_l^2)\in L_{I_s}}U_l^{a_r}+\sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}}U_l^{a_r}\Bigr]\le z_r\Bigr\}\\
&\subseteq\Bigl\{(\sigma(a_r)P_{a_r}^n)^{-1}\Bigl|\sum_{(j_l^1,j_l^2)\in L_{I_s}}U_l^{a_r}\Bigr|\ge\Gamma_p\Bigr\}\bigcup\Bigl\{(\sigma(a_r)P_{a_r}^n)^{-1}\sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}}U_l^{a_r}\le\Gamma_p+z_r\Bigr\},
\end{aligned}$$

and

$$\begin{aligned}
&\Bigl\{(\sigma(a_r)P_{a_r}^n)^{-1}\sum_{(j_l^1,j_l^2)\in L\backslash L_{I_s}}U_l^{a_r}\le z_r-\Gamma_p\Bigr\}\\
&\subseteq\Bigl\{(\sigma(a_r)P_{a_r}^n)^{-1}\Bigl|\sum_{(j_l^1,j_l^2)\in L_{I_s}}U_l^{a_r}\Bigr|\ge\Gamma_p\Bigr\}\bigcup\Bigl\{(\sigma(a_r)P_{a_r}^n)^{-1}\sum_{v=1}^{q}U_v^{a_r}\le z_r\Bigr\}.
\end{aligned}$$

Therefore similarly to (B.108) and (B.110), we know

$$\text{(B.116)} \le P_{ys}P_{+z} + \sum_{r=1}^{m} P_{s_r}, \quad \text{(B.116)} \ge P_{ys}P_{-z} - \Big(\sum_{r=1}^{m} P_{s_r}\Big), \tag{B.117}$$

where P_{ys} is defined in (B.109), and we further define

$$\begin{aligned} P_{+z} &= P\Big(\bigcap_{r=1}^{m}\Big\{(\sigma(a_r)P_{a_r}^{n})^{-1} \sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} \le z_r + \Gamma_p\Big\}\Big), \\ P_{s_r} &= P\Big((\sigma(a_r)P_{a_r}^{n})^{-1}\Big|\sum_{(j_l^1,j_l^2)\in L_{I_s}} U_l^{a_r}\Big| \ge \Gamma_p\Big), \\ P_{-z} &= P\Big(\bigcap_{r=1}^{m}\Big\{(\sigma(a_r)P_{a_r}^{n})^{-1} \sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} \le z_r - \Gamma_p\Big\}\Big). \end{aligned}$$

We note that the cardinality of L_{I_s} is no greater than $2ps$, which is $o(p^2)$. Similarly to Section B.5.8, we know $(\sigma(a_r)P_{a_r}^{n})^{-1} \times \sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} \xrightarrow{P} 0$ for $r = 1,\ldots,m$. Combined with Theorem 3.2.1, we know $\{(\sigma(a_r)P_{a_r}^{n})^{-1} \times \sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} : r = 1,\ldots,m\}$ converges to $\mathcal{N}(0, I_m)$ and thus are asymptotically independent. We then have

$$\Big|P_{+z} - \prod_{r=1}^{m} P_{+z_r}\Big| \to 0, \quad \Big|P_{-z} - \prod_{r=1}^{m} P_{-z_r}\Big| \to 0, \tag{B.118}$$

where we define

$$\begin{aligned} P_{+z_r} =& P\Big((\sigma(a_r)P_{a_r}^{n})^{-1} \sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} \le z_r + \Gamma_p\Big), \\ P_{-z_r} =& P\Big((\sigma(a_r)P_{a_r}^{n})^{-1} \sum_{(j_l^1,j_l^2)\in L\setminus L_{I_s}} U_l^{a_r} \le z_r - \Gamma_p\Big). \end{aligned}$$

Similarly to (B.111), for each $r = 1,\ldots,m$, we have

$$|P_{+z_r} - P_{z_r}| \le C\Gamma_p \quad \text{and} \quad |P_{-z_r} - P_{z_r}| \le C\Gamma_p, \tag{B.119}$$

where we define $P_{z_r} = P((\sigma(a_r)P_{a_r}^n)^{-1}\sum_{v=1}^{q} U_v^{a_r} \leq z_r)$. Combining (B.118) and (B.119), we have

$$\Big|P_{+z} - \prod_{r=1}^{m} P_{z_r}\Big| \to 0 \quad \text{and} \quad \Big|P_{-z} - \prod_{r=1}^{m} P_{z_r}\Big| \to 0.$$

By (B.117) and (B.119),

$$\Big|(\text{B.116}) - P_{ys}\prod_{r=1}^{m} P_{z_r}\Big| \leq o(1)P_{ys} + \sum_{r=1}^{m} P_{s_r}. \tag{B.120}$$

Given (B.120), similarly to (B.112), we have

$$\begin{aligned}
&P(\cup_{l=1}^{q} E_l) \\
&\leq \sum_{s=1}^{d-1}(-1)^{s-1}\sum_{1\leq l_1<\ldots<l_s\leq q}\Big\{P_{ys}\prod_{r=1}^{m}P_{z_r} + (-1)^{s-1}\times\Big[o(1)P_{ys} + \sum_{r=1}^{m}P_{s_r}\Big]\Big\} \\
&\leq H_{d-1}\prod_{r=1}^{m}P_{z_r} + \sum_{s=1}^{d-1}\sum_{1\leq l_1<\ldots<l_s\leq q}\Big\{o(1)P_{ys} + \sum_{r=1}^{m}P_{s_r}\Big\} \\
&\leq P_y\prod_{r=1}^{m}P_{z_r} + |H_{d-1} - H_d|\prod_{r=1}^{m}P_{z_r} + \sum_{s=1}^{d-1}\sum_{1\leq l_1<\ldots<l_s\leq q}\Big\{o(1)P_{ys} + \sum_{r=1}^{m}P_{s_r}\Big\},
\end{aligned}$$

where H_{d-1} follows the definition in (B.113) and we use (B.104) and the definition (B.114) in the last inequality. By Lemma B.5.5, $|H_{d-1} - H_d| \to 0$; by Lemma B.5.6, $o(1)\sum_{s=1}^{d-1}\sum_{1\leq l_1<\ldots<l_s\leq q} P_{ys} \to 0$; by Lemma B.5.7, $\sum_{r=1}^{m}\sum_{s=1}^{d-1}\sum_{1\leq l_1<\ldots<l_s\leq q} P_{s_r} =\to 0$.

In summary, we have shown that $P(\cup_{l=1}^{q} E_l) \leq P_y \times \prod_{r=1}^{m} P_{z_r} + o(1)$. Moreover, following similar arguments, we have $P(\cup_{l=1}^{q} E_l) \geq P_y \times \prod_{r=1}^{m} P_{z_r} + o(1)$. Therefore, $|P(\cup_{l=1}^{q} E_l) - P_y \times \prod_{r=1}^{m} P_{z_r}| \to 0$ is obtained, that is,

$$\Big|P(\cup_{l=1}^{q} E_l) - P(\cup_{l=1}^{q}\{(\hat{G}_l)^2 > ny_p\})\prod_{r=1}^{m} P((\sigma(a_r)P_{a_r}^n)^{-1}\sum_{v=1}^{q} U_v^{a_r} \leq z_r)\Big| \to 0.$$

Since (B.115) $= P(\cup_{l=1}^{q} E_l)$, $\{\hat{M}_n/n > y_p\} = \cup_{l=1}^{q}\{(\hat{G}_l)^2 > ny_p\}$ and $\tilde{\mathcal{U}}(a_r) = 2(P_{a_r}^n)^{-1}\sum_{v=1}^{q} U_v^{a_r}$, we know Lemma B.1.7 is proved for $m > 1$.

B.5.9 Proof of Lemma B.1.8

Similarly to Section B.5.8, we prove Lemma B.1.8 for $m = 1$ first and then discuss $m > 1$.

Proof for $m = 1$ Specifically, in this section, we prove for finite integer a,

$$\left| P\Big(\frac{M_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \le z\Big) - P\Big(\frac{M_n}{n} > y_p\Big) P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \le z\Big) \right| \to 0. \tag{B.121}$$

To prove (B.121), we start by proving the following two conclusions (B.122) and (B.123), which suggest that M_n and $\hat{M}_n$ have small difference in probability. To be specific, as $n, p \to \infty$,

$$|P(M_n/n > y_p) - P(\hat{M}_n/n > y_p)| \to 0, \tag{B.122}$$

and

$$|P(M_n/n > y_p\,, \tilde{\mathcal{U}}(a)/\sigma(a) \le z) - P(\hat{M}_n/n > y_p\,, \tilde{\mathcal{U}}(a)/\sigma(a) \le z)| \to 0. \tag{B.123}$$

To prove (B.122) and (B.123), recall that in (B.7), M_n and $\hat{M}_n$ are defined using $\tilde{G}_l$ and $\hat{G}_l$ respectively. We next focus on the difference between $\tilde{G}_l$ and $\hat{G}_l$. Since $\tilde{G}_l$ and $\hat{G}_l$ will not change if the data $x_{i,j}$ is scaled by its standard deviation, then we assume, without loss of generality, $\sigma_{j,j} = 1$, $j = 1, \ldots, p$ in the following discussion.

By the definitions in (B.7), we have

$$P\Big(\max_{1\le l\le q} |\tilde{G}_l - \hat{G}_l| \ge (\log p)^{-1}\Big) \le P\Big(\max_{1\le l\le q}\max_{1\le i\le n} |x_{i,j_l^1} x_{i,j_l^2}| \ge \tau_n\Big).$$

Note that $|x_{i,j_l^1} x_{i,j_l^2}| \le (x^2_{i,j_l^1} + x^2_{i,j_l^2})/2$. Then

$$
\begin{aligned}
& P\Big(\max_{1\le l\le q}\max_{1\le i\le n} |x_{i,j_l^1} x_{i,j_l^2}| \ge \tau_n\Big) \\
\le\ & P\Big(\max_{1\le l\le q}\max_{1\le i\le n} (x^2_{i,j_l^1} + x^2_{i,j_l^2}) \ge 2\tau_n\Big) \\
\le\ & P\Big(\max_{1\le l\le q}\max_{1\le i\le n} x^2_{i,j_l^1} \ge \tau_n\Big) + P\Big(\max_{1\le l\le q}\max_{1\le i\le n} x^2_{i,j_l^2} \ge \tau_n\Big) && \text{(B.124)} \\
\le\ & 2P\Big(\max_{1\le j\le p}\max_{1\le i\le n} x^2_{i,j} \ge \tau_n\Big) && \text{(B.125)} \\
\le\ & 2np \max_{1\le j\le p} P(|x^2_{1,j}| \ge \tau_n).
\end{aligned}
$$

From (B.124) to (B.125), we use $\max_{1\le l\le q} x^2_{i,j_l^k} = \max_{1\le j\le p} x^2_{i,j}$ for each i and $k = 1, 2$. To see this, recall the notation defined in Section B.1.2 (on Page 256). In particular, subscript l is defined to indicate a pair of indexes (j_l^1, j_l^2) with $1 \le j_l^1 < j_l^2 \le p$. Since j_l^1 and j_l^1 only take values from the range $\{1, \dots, p\}$, we know $\{j_l^k : 1 \le l \le q\} \subseteq \{1, \dots, p\}$ for $k = 1, 2$, and then $\max_{1\le l\le q} x^2_{i,j_l^1} = \max_{1\le j\le p} x^2_{i,j}$. Moreover, by Condition 3.2.3 with $\varsigma = 2$,

$$
np \max_{1\le j\le p} P(|x^2_{1,j}| \ge \tau_n) \le Cnp(n+p)^{-\tau}\mathrm{E}\exp(x^2_{1,1}) \to 0.
$$

It follows that $P(\max_{1\le l\le q} |\tilde G_l - \hat G_l| \ge (\log p)^{-1}) \to 0$. Conditioning on $\max_{1\le l\le q} |\tilde G_l - \hat G_l| \le (\log p)^{-1}$, by Lemma B.5.3 and $|\hat G_l| \le \tau_n$,

$$
\begin{aligned}
|M_n - \hat M_n| &= \Big|\max_{1\le l\le q}(\tilde G_l)^2 - \max_{1\le l\le q}(\hat G_l)^2\Big| \\
&\le 2\max_{1\le l\le q}|\hat G_l| \max_{1\le l\le q}|\tilde G_l - \hat G_l| + \max_{1\le l\le q}|\tilde G_l - \hat G_l|^2 \\
&\le 2\tau_n/\log p + (\log p)^{-2}.
\end{aligned}
$$

Recall that $\tau_n = O(\log(p+n))$, then $|M_n/n - \hat M_n/n| \xrightarrow{P} 0$. Therefore (B.122) and (B.123) are obtained.

Given (B.122) and (B.123), we next prove (B.121). In particular, we write

$$P\Big(\frac{M_n}{n} > y_p, \frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \le z\Big) - P\Big(\frac{M_n}{n} > y_p\Big)P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} \le z\Big) = \Delta_{P,1} + \Delta_{P,2} + \Delta_{P,3},$$

where we define

$$\begin{aligned}
\Delta_{P,1} &= P\Big(M_n/n > y_p, \tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big) - P\Big(\hat{M}_n/n > y_p, \tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big),\\
\Delta_{P,2} &= P\Big(\hat{M}_n/n > y_p, \tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big) - P\Big(\hat{M}_n/n > y_p\Big) \times P\Big(\tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big),\\
\Delta_{P,3} &= P\Big(\hat{M}_n/n > y_p\Big) \times P\Big(\tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big)\\
&\qquad - P\Big(M_n/n > y_p\Big) \times P\Big(\tilde{\mathcal{U}}(a)/\sigma(a) \le z\Big).
\end{aligned}$$

Note that the left hand side of (B.121) $\le |\Delta_{p,1}| + |\Delta_{p,2}| + |\Delta_{p,3}|$. By Lemma B.1.7, $|\Delta_{p,2}| \to 0$; by (B.123), $|\Delta_{p,1}| \to 0$; by $|\Delta_{p,3}| \le |P(\hat{M}_n/n > y_p) - P(M_n/n > y_p)|$ and (B.122), $|\Delta_{p,3}| \to 0$. In summary, (B.121) is proved.

Proof for $m > 1$ Following the proof in Section B.5.9, we know that (B.122) still holds and similarly to (B.123),

$$\begin{aligned}
&|P(M_n/n > y_p, \tilde{\mathcal{U}}(a_1)/\sigma(a_1) \le z_1, \dots, \tilde{\mathcal{U}}(a_m)/\sigma(a_m) \le z_m)\\
&\quad - P(\hat{M}_n/n > y_p, \tilde{\mathcal{U}}(a_1)/\sigma(a_1) \le z_1, \dots, \tilde{\mathcal{U}}(a_m)/\sigma(a_m) \le z_m)| \to 0.
\end{aligned}$$

Given these results and Lemma B.1.7, we know that Lemma B.1.8 holds for $m > 1$, following the arguments in Section B.5.9 similarly.

B.5.10 Proof of Lemma B.1.9

Similarly to Section B.5.9, we first prove Lemma B.1.9 for $m = 1$ in Section B.5.10, and then discuss the case for $m > 1$ in Section B.5.10.

Proof for $m = 1$ Specifically, in this section, we prove for finite integer a and given z,

$$\Big|P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} \le z,\, n\mathcal{U}^2(\infty) > y_p\Big) - P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} \le z\Big)P\Big(n\mathcal{U}^2(\infty) > y_p\Big)\Big| \to 0. \tag{B.126}$$

To prove this, we use M_n/n as an intermediate variable and first show

$$\Big|P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} > z, \frac{M_n}{n} > y_p\Big) - P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} > z\Big)P\Big(\frac{M_n}{n} > y_p\Big)\Big| \to 0. \tag{B.127}$$

To facilitate the proof, we define some notation. Given small constant $\epsilon > 0$,

$$\begin{aligned}
P_{uz} &= P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} > z\Big), \quad P_{zy} = P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} > z, \frac{M_n}{n} > y_p\Big),\\
P_{uz+\epsilon} &= P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} > z+\epsilon\Big), \quad P_{z+\epsilon} = P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} > z+\epsilon, \frac{M_n}{n} > y_p\Big),\\
P_{uz-\epsilon} &= P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} > z-\epsilon\Big), \quad P_{z-\epsilon} = P\Big(\frac{\tilde{\mathcal{U}}(a)}{\sigma(a)} > z-\epsilon, \frac{M_n}{n} > y_p\Big),\\
P_{y_p} &= P\Big(\frac{M_n}{n} > y_p\Big),
\end{aligned}$$

$\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution, and $\bar{\Phi}(\cdot) = 1 - \Phi(\cdot)$. Then

$$\text{(B.127)} = |P_{zy} - P_{uz} \times P_{y_p}| \le |P_{zy} - P_{z+\epsilon}| + |P_{z+\epsilon} - P_{uz+\epsilon}P_{y_p}| + |P_{uz+\epsilon}P_{y_p} - P_{uz}P_{y_p}|.$$

We next show (B.127) $\to 0$ by proving the three parts above all converges to 0 respectively.

First we show $|P_{zy} - P_{z+\epsilon}| \to 0$. Note that $P_{z+\epsilon} \le P_{zy} \le P_{z-\epsilon}$, then $|P_{zy} - P_{z+\epsilon}| \le$

$|P_{z-\epsilon} - P_{z+\epsilon}|$. In addition,

$$
\begin{aligned}
& |P_{z-\epsilon} - P_{z+\epsilon}| \\
\leq\ & |P_{z-\epsilon} - P_{uz-\epsilon} \times P_{y_p}| + |P_{uz-\epsilon} \times P_{y_p} - P_{uz+\epsilon} \times P_{y_p}| + |P_{uz+\epsilon} \times P_{y_p} - P_{z+\epsilon}| \\
\leq\ & o(1) + |P_{uz+\epsilon} - P_{uz-\epsilon}|,
\end{aligned}
$$

where we use (B.121) in the last inequality. Moreover, by the proof of Theorem 3.2.1 in Section B.1.1, we know $\tilde{\mathcal{U}}(a)/\sigma(a) \xrightarrow{D} \mathcal{N}(0,1)$. Thus when $n, p \to \infty$ and $\epsilon \to 0$,

$$
\begin{aligned}
& |P_{uz+\epsilon} - P_{uz-\epsilon}| \\
\leq\ & |P_{uz+\epsilon} - \bar{\Phi}(z+\epsilon)| + |\bar{\Phi}(z+\epsilon) - \bar{\Phi}(z-\epsilon)| + |P_{uz-\epsilon} - \bar{\Phi}(z-\epsilon)| + o(1) \\
\to\ & 0.
\end{aligned}
$$

Second, we know $|P_{z+\epsilon} - P_{uz+\epsilon}P_{y_p}| \to 0$ by (B.121). Last, we show $|P_{uz+\epsilon}P_{y_p} - P_{uz}P_{y_p}| \to 0$. By the proof of Theorem 3.2.1 in Section B.1.1, we know $\tilde{\mathcal{U}}(a)/\sigma(a) \xrightarrow{D} \mathcal{N}(0,1)$, $\{\mathcal{U}(a) - \tilde{\mathcal{U}}(a)/\sigma(a)\} \xrightarrow{P} 0$, and $\mathcal{U}(a)/\sigma(a) \xrightarrow{D} \mathcal{N}(0,1)$. Thus when $n, p \to \infty$ and $\epsilon \to 0$,

$$
\begin{aligned}
& |P_{uz+\epsilon}P_{y_p} - P_{uz}P_{y_p}| \\
\leq\ & |P_{uz+\epsilon} - P_{uz}| \\
\leq\ & |P_{uz+\epsilon} - \bar{\Phi}(z+\epsilon)| + |\bar{\Phi}(z+\epsilon) - \bar{\Phi}(z)| + |P_{uz} - \bar{\Phi}(z)| + o(1) \\
\to\ & 0.
\end{aligned}
$$

In summary (B.127) is proved.

We next prove (B.126) similarly to the proof of (B.127). Specifically, we write

$$\begin{aligned}&\left|P\Big(n\mathcal{U}^2(\infty) > y_p, \frac{\mathcal{U}(a)}{\sigma(a)} \leq z\Big) - P\Big(n\mathcal{U}^2(\infty) > y_p\Big)P\Big(\frac{\mathcal{U}(a)}{\sigma(a)} \leq z\Big)\right|\\ = &\,|P_{z0} - P_{y0} \times P_{uz}|,\end{aligned}$$

where we define $P_{z0} = P(n\mathcal{U}^2(\infty) > y_p, \frac{\mathcal{U}(a)}{\sigma(a)} > z)$ and $P_{y0} = P(n\mathcal{U}^2(\infty) > y_p)$. Note that

$$|P_{z0} - P_{y0}P_{uz}| \leq |P_{z0} - P_{zy-\epsilon}| + |P_{zy-\epsilon} - P_{y-\epsilon}P_{uz}| + |P_{y-\epsilon}P_{uz} - P_{y0}P_{uz}|,$$

where

$$\begin{aligned}P_{zy-\epsilon} &= P\Big(\frac{M_n}{n} > y_p - \epsilon, \frac{\mathcal{U}(a)}{\sigma(a)} > z\Big), \quad P_{y-\epsilon} = P\Big(\frac{M_n}{n} > y_p - \epsilon\Big),\\ P_{zy+\epsilon} &= P\Big(\frac{M_n}{n} > y_p + \epsilon, \frac{\mathcal{U}(a)}{\sigma(a)} > z\Big), \quad P_{y+\epsilon} = P\Big(\frac{M_n}{n} > y_p + \epsilon\Big).\end{aligned}$$

To prove (B.126), we will show $|P_{z0} - P_{zy-\epsilon}|$, $|P_{zy-\epsilon} - P_{y-\epsilon}P_{uz}|$, and $|P_{y-\epsilon}P_{uz} - P_{y0}P_{uz}|$ all converge to 0 respectively.

First we show $|P_{z0} - P_{zy-\epsilon}| \to 0$. Note that $W_n \xrightarrow{P} 0$ where $W_n = (n^2\mathcal{U}^2(\infty) - M_n)/n$ by the proof of Theorem 3 in Cai and Jiang (2011). Then for any $\epsilon > 0$, $P(|W_n| > \epsilon) \to 0$. Since $P_{zy+\epsilon} - P(|W_n| > \epsilon) \leq P_{z0} \leq P_{zy-\epsilon} + P(|W_n| > \epsilon)$, we have $|P_{z0} - P_{zy-\epsilon}| \leq |P_{zy-\epsilon} - P_{zy+\epsilon}| + o(1)$. Furthermore,

$$\begin{aligned}&|P_{zy-\epsilon} - P_{zy+\epsilon}|\\ \leq &\,|P_{zy-\epsilon} - P_{y-\epsilon}P_{uz}| + |P_{y-\epsilon}P_{uz} - P_{y+\epsilon}P_{uz}| + |P_{y+\epsilon}P_{uz} - P_{zy+\epsilon}| \to 0,\end{aligned}$$

where the last equation follows from (B.127) and $|P_{y-\epsilon} - P_{y+\epsilon}| \to 0$ when $\epsilon \to 0$. Second we know $|P_{zy-\epsilon} - P_{y-\epsilon}P_{uz}| \to 0$ by (B.127). Last we show $|P_{y-\epsilon}P_{uz} - P_{y0}P_{uz}| \to 0$. In particular, as $P_{y+\epsilon} - P(|W_n| > \epsilon) \leq P_{y0} \leq P_{y-\epsilon} + P(|W_n| > \epsilon)$ and $P(|W_n| >$

$\epsilon) \to 0$, we have

$$|P_{y-\epsilon}P_{uz} - P_{y0}P_{uz}| \le |P_{y-\epsilon} - P_{y0}| \le |P_{y-\epsilon} - P_{y+\epsilon}| + o(1) \to 0.$$

In summary, Lemma B.1.9 is proved.

Proof for $m > 1$ Note that $W_n = \{n^2\mathcal{U}^2(\infty) - M_n\}/n \xrightarrow{P} 0$ and $\tilde{\mathcal{U}}^*(a_r) = \mathcal{U}(a_r) - \tilde{\mathcal{U}}(a_r) \xrightarrow{P} 0$ for each $r = 1, \ldots, m$ as argued in Section B.1.2. Therefore when m is finite, the arguments above can be applied to prove Lemma B.1.9 for $m > 1$ similarly.

B.5.11 Proof of Lemma B.1.10

We first prove $\mathbb{V}_{u,1}(a)/\mathrm{E}\{\mathbb{V}_{u,1}(a)\} \xrightarrow{P} 1$. To prove this, and it suffices to show $\mathrm{var}\{\mathbb{V}_{u,1}(a)\}/\mathrm{E}^2\{\mathbb{V}_{u,1}(a)\} \to 0$. By the notation defined in Section B.5.1, we have

$$\begin{aligned}
&\mathrm{var}\{\mathbb{V}_{u,1}(a)\} \\
&= \mathrm{E}\{\mathbb{V}_{u,1}^2(a)\} - \mathrm{E}^2\{\mathbb{V}_{u,1}(a)\} \\
&= \frac{(2a!)^2}{(P_a^n)^4} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a);\\ 1\le j_1\ne j_2\le p,\\ 1\le j_3\ne j_4\le p}} \Big[\mathrm{E}\Big(\prod_{t=1}^{a} x_{i_t,j_1}^2 x_{i_t,j_2}^2 x_{\tilde{i}_t,j_3}^2 x_{\tilde{i}_t,j_4}^2\Big) - \Big\{\mathrm{E}(x_{1,j_1}^2 x_{1,j_2}^2)\mathrm{E}(x_{1,j_3}^2 x_{1,j_4}^2)\Big\}^a\Big].
\end{aligned}$$

To evaluate $\mathrm{var}\{\mathbb{V}_{u,1}(a)\}$, we consider the summed term in $\mathrm{var}\{\mathbb{V}_{u,1}(a)\}$, that is,

$$\mathrm{E}\Big(\prod_{t=1}^{a} x_{i_t,j_1}^2 x_{i_t,j_2}^2 x_{\tilde{i}_t,j_3}^2 x_{\tilde{i}_t,j_4}^2\Big) - \{\mathrm{E}(x_{1,j_1}^2 x_{1,j_2}^2)\}^a\{\mathrm{E}(x_{1,j_3}^2 x_{1,j_4}^2)\}^a. \tag{B.128}$$

When $\{\mathbf{i}\} \cap \{\tilde{\mathbf{i}}\} = \emptyset$, (B.128) $= 0$. We then know that (B.128) $\ne 0$ only when $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \le 2a - 1$. Along with Condition 3.2.1, we have

$$|\mathrm{var}\{\mathbb{V}_{u,1}(a)\}| \le Cp^4 n^{-4a} n^{2a-1},$$

which induces $\text{var}\{\mathbb{V}_{u,1}(a)\} = O(p^4 n^{-2a-1})$. By (B.55) and (B.60), we know $\text{E}\{\mathbb{V}_{u,1}(a)\} = \Theta(p^2 n^{-a})$. It follows that $\text{var}\{\mathbb{V}_{c,1}(a)\}/\text{E}^2\{\mathbb{V}_{c,1}(a)\} \to 0$ as $n \to \infty$.

We next prove $\mathbb{V}_{u,2}(a)/\text{E}\{\mathbb{V}_{u,1}(a)\} \xrightarrow{P} 0$. By the Markov's inequality, it suffices to prove $\text{E}\{\mathbb{V}_{u,2}^2(a)\} = o(1)[\text{E}\{\mathbb{V}_{u,1}(a)\}]^2$. As $\text{E}\{\mathbb{V}_{u,1}(a)\} = \Theta(p^2 n^{-a})$, it is sufficient to prove $\text{E}\{\mathbb{V}_{u,2}^2(a)\} = o(p^4 n^{-2a})$ below.

We first derive the form of $\mathbb{V}_{u,2}(a)$. In particular, when $a = 1$,

$$
\begin{aligned}
\mathbb{V}_{u,2}(1) &= \mathbb{V}_u(1) - \mathbb{V}_{u,1}(1) \\
&= \frac{1}{n^2} \sum_{1 \le j_1 \ne j_2 \le p} \sum_{i \in \mathcal{P}(n,1)} \Big\{ (x_{i,j_1} - \bar{x}_{j_1})^2 (x_{i,j_2} - \bar{x}_{j_2})^2 - x_{i,j_1}^2 x_{i,j_2}^2 \Big\} \\
&= \frac{1}{n^2} \sum_{1 \le j_1 \ne j_2 \le p} \sum_{1 \le i \le n} \sum_{\substack{s_1 + r_1 = 1, \\ s_2 + r_2 = 1}} C_{s_1,r_1,s_2,r_2} \prod_{k=1}^{2} \Big\{ (-x_{i,j_k} \bar{x}_{j_k})^{s_k} (\bar{x}_{j_k}^2)^{r_k} \Big\},
\end{aligned}
$$

where C_{s_1,r_1,s_2,r_2} is some constant and we use

$$
\begin{aligned}
& (x_{i,j_1} - \bar{x}_{i,j_1})^2 (x_{i,j_2} - \bar{x}_{i,j_2})^2 - x_{i,j_1}^2 x_{i,j_2}^2 \\
= {} & (x_{i,j_1}^2 - 2x_{i,j_1}\bar{x}_{j_1} + \bar{x}_{j_1}^2)(x_{i,j_2}^2 - 2x_{i,j_2}\bar{x}_{j_2} + \bar{x}_{j_2}^2) - x_{i,j_1}^2 x_{i,j_2}^2 \\
= {} & \sum_{s_1 + r_1 = 1,\, s_2 + r_2 = 1} \Big\{ (-2x_{i,j_1}\bar{x}_{j_1})^{s_1} (\bar{x}_{j_1}^2)^{r_1} \Big\} \times \Big\{ (-2x_{i,j_2}\bar{x}_{j_2})^{s_2} (\bar{x}_{j_2}^2)^{r_2} \Big\}.
\end{aligned}
$$

Following this example, we similarly give the form of $\mathbb{V}_{u,2}(a)$ for general $a \ge 1$. Given tuple $\mathbf{i} \in \mathcal{P}(n,a)$, for $k = 1, 2$, let $\mathbf{i}_{(a-r_k)}^{(k)}$ represent a sub-tuple of $\mathbf{i}$ with length $a - r_k$, and define $\mathcal{S}(\mathbf{i}, a - r_k)$ to be the collection of sub-tuples of $\mathbf{i}$ with length $a - r_k$. Then for $a \ge 1$, we write $\mathbb{V}_{u,2}(a) = \sum_{1 \le s_1 + r_1 \le a, 1 \le s_2 + r_2 \le a} T_{s_1,r_1,s_2,r_2}$, where

$$
\begin{aligned}
T_{s_1,r_1,s_2,r_2} &= \frac{a!}{(P_a^n)^2} \sum_{1 \le j_1 \ne j_2 \le p} \sum_{\substack{\mathbf{i} \in \mathcal{P}(n,a); \\ \mathbf{i}_{(a-r_k)}^{(k)} \in \mathcal{S}(\mathbf{i}, a - r_k):\, k = 1,2}} C_{s_1,r_1,s_2,r_2} \\
&\quad \times \prod_{k=1}^{2} \Big\{ (-\bar{x}_{j_k})^{s_k + 2r_k} \prod_{t_k=1}^{s_k} x_{i_{t_k}^{(k)}, j_k} \prod_{t_k = s_k + 1}^{a - r_k} (x_{i_{t_k}^{(k)}, j_k})^2 \Big\}.
\end{aligned}
$$

When a is finite, it suffices to prove $\mathrm{E}(T^2_{s_1,r_1,s_2,r_2}) = o(p^4 n^{-2a})$. Note that

$$
\begin{aligned}
& \mathrm{E}(T^2_{s_1,r_1,s_2,r_2}) \\
= \; & \frac{(a!)^2}{(P^n_a)^4} \sum_{\substack{1\le j_1\ne j_2\le p \\ 1\le \tilde{j}_1\ne \tilde{j}_2\le p}} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a); \\ \mathbf{i}^{(k)}_{(a-r_k)}\in\mathcal{S}(\mathbf{i},a-r_k):\, k=1,2; \\ \tilde{\mathbf{i}}^{(k)}_{(a-r_k)}\in\mathcal{S}(\tilde{\mathbf{i}},a-r_k):\, k=1,2}} C^2_{s_1,r_1,s_2,r_2} \\
& \times \mathrm{E}\Big\{ \prod_{k=1}^{2} (\bar{x}_{j_k}\bar{x}_{\tilde{j}_k})^{s_k+2r_k} \prod_{t_k=1}^{s_k} (x_{i^{(k)}_{t_k},j_k} x_{\tilde{i}^{(k)}_{t_k},\tilde{j}_k}) \prod_{t_k=s_k+1}^{a-r_k} (x_{i^{(k)}_{t_k},j_k} x_{\tilde{i}^{(k)}_{t_k},\tilde{j}_k})^2 \Big\}.
\end{aligned}
$$

Recall that $\bar{x}_j = \sum_{i=1}^n x_{i,j}/n$. We have

$$
\begin{aligned}
& \mathrm{E}(T^2_{s_1,r_1,s_2,r_2}) \\
= \; & \frac{(a!)^2}{(P^n_a)^4 n^{\sum_{k=1}^2 (2s_k+4r_k)}} \sum_{\substack{1\le j_1\ne j_2\le p \\ 1\le \tilde{j}_1\ne \tilde{j}_2\le p}} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a); \\ \mathbf{i}^{(k)}_{(a-r_k)}\in\mathcal{S}(\mathbf{i},a-r_k):\, k=1,2; \\ \tilde{\mathbf{i}}^{(k)}_{(a-r_k)}\in\mathcal{S}(\tilde{\mathbf{i}},a-r_k):\, k=1,2}} C^2_{s_1,r_1,s_2,r_2} \\
& \times \sum_{\mathbf{m}^{(k)},\tilde{\mathbf{m}}^{(k)}\in\mathcal{C}(n,s_k+2r_k);\, k=1,2} T\{\mathbf{i}^{(k)}_{(a-r_k)}, \tilde{\mathbf{i}}^{(k)}_{(a-r_k)}, \mathbf{m}^{(k)}, \tilde{\mathbf{m}}^{(k)}; k=1,2\},
\end{aligned}
$$

where $\mathcal{C}(n, s_k+2r_k)$ follows the notation in Section B.5.1 and

$$
\begin{aligned}
& T\{\mathbf{i}^{(k)}_{(a-r_k)}, \tilde{\mathbf{i}}^{(k)}_{(a-r_k)}, \mathbf{m}^{(k)}, \tilde{\mathbf{m}}^{(k)}; k=1,2\} \\
= \; & \mathrm{E}\Big\{ \prod_{k=1}^{2} \prod_{\tilde{t}_k=1}^{s_k+2r_k} (x_{m_{\tilde{t}_k},j_k} x_{\tilde{m}_{\tilde{t}_k},\tilde{j}_k}) \prod_{t_k=1}^{s_k} (x_{i^{(k)}_{t_k},j_k} x_{\tilde{i}^{(k)}_{t_k},\tilde{j}_k}) \prod_{t_k=s_k+1}^{a-r_k} (x_{i^{(k)}_{t_k},j_k} x_{\tilde{i}^{(k)}_{t_k},\tilde{j}_k})^2 \Big\}.
\end{aligned}
$$

Since $\mathrm{E}(x_{i,j}) = 0$, $T\{\mathbf{i}^{(k)}_{(a-r_k)}, \tilde{\mathbf{i}}^{(k)}_{(a-r_k)}, \mathbf{m}^{(k)}, \tilde{\mathbf{m}}^{(k)}; k=1,2\} \neq 0$ only when

$$
\left| \bigcup_{k=1}^{2} \{\mathbf{m}^{(k)}\} \cup \{\tilde{\mathbf{m}}^{(k)}\} \cup \{\mathbf{i}^{(k)}_{(a-r_k)}\} \cup \{\tilde{\mathbf{i}}^{(k)}_{(a-r_k)}\} \right| - \left| \bigcup_{k=1}^{2} \{\mathbf{i}^{(k)}_{(a-r_k)}\} \cup \{\tilde{\mathbf{i}}^{(k)}_{(a-r_k)}\} \right| \le \sum_{k=1}^{2} (s_k + 2r_k).
$$

Since $\mathbf{i}^{(k)}_{(a-r_k)}$ and $\tilde{\mathbf{i}}^{(k)}_{(a-r_k)}$ are sub-tuples of $\mathbf{i}$ and $\tilde{\mathbf{i}} \in \mathcal{P}(n,a)$, $|\cup_{k=1}^2 \{\mathbf{i}^{(k)}_{(a-r_k)}\} \cup \{\tilde{\mathbf{i}}^{(k)}_{(a-r_k)}\}| \le$

$|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \leq 2a$. Therefore,

$$\left| \bigcup_{k=1}^{2} \{\mathbf{m}^{(k)}\} \cup \{\tilde{\mathbf{m}}^{(k)}\} \cup \{\mathbf{i}^{(k)}_{(a-r_k)}\} \cup \{\tilde{\mathbf{i}}^{(k)}_{(a-r_k)}\} \right| \leq 2a + \sum_{k=1}^{2} (s_k + 2r_k). \tag{B.129}$$

By (B.129) and the boundedness of moments in Condition 3.2.4, we have

$$\begin{aligned} \mathrm{E}(T^2_{s_1,r_1,s_2,r_2}) &= O\Big(p^4 n^{-4a - \sum_{k=1}^{2}(2s_k+4r_k) + 2a + \sum_{k=1}^{2}(s_k+2r_k)}\Big) \\ &= O(p^4 n^{-2a-\sum_{k=1}^{2}(s_k+2r_k)}) = o(p^4 n^{-2a}), \end{aligned}$$

where we use $\sum_{k=1}^{2}(s_k + 2r_k) \geq 1$.

B.5.12 Proof of Lemma B.1.11

To show $\mathrm{var}\{\mathcal{U}(a)\} \simeq \mathrm{var}(T_{U,a,1,1})$, it suffices to prove $\mathrm{var}(T_{U,a,1,1}) = \Theta(p^2 n^{-a})$, $\mathrm{var}(T_{U,a,1,2}) = o(p^2 n^{-a})$ and $\mathrm{var}(T_{U,a,2}) = o(p^2 n^{-a})$. The following three sections B.5.12.1–B.5.12.3 prove the three results respectively.

B.5.12.1 Proof of $\mathrm{var}(T_{U,a,1,1}) = \Theta(p^2 n^{-a})$

As $\mathrm{E}(T_{U,a,1,1}) = 0$, $\mathrm{var}(T_{U,a,1,1}) = \mathrm{E}(T^2_{U,a,1,1})$, and we have

$$\mathrm{var}(T_{U,a,1,1}) = \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} (P_a^n)^{-2} \sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a)} \mathrm{E}\Big(\prod_{k=1}^{a} x_{i_k,j_1} x_{i_k,j_2} x_{\tilde{i}_k,j_3} x_{\tilde{i}_k,j_4}\Big).$$

Similarly to Section B.5.2, $\mathrm{E}(\prod_{k=1}^{a} x_{i_k,j_1} x_{i_k,j_2} x_{\tilde{i}_k,j_3} x_{\tilde{i}_k,j_4}) \neq 0$ only when $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$. Therefore, $\mathrm{var}(T_{U,a,1,1}) = \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} (P_a^n)^{-1} a! \times \{\mathrm{E}(\prod_{t=1}^{4} x_{1,j_t})\}^a$. By Condition 3.2.6, as $(j_1,j_2),(j_3,j_4) \in J_A^c$,

$$\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) = \kappa_1(\sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3}). \tag{B.130}$$

We next evaluate (B.130) by discussing three cases on (j_1, j_2, j_3, j_4). First, if $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 2$, (B.130) $= \kappa_1 \sigma_{j_1,j_1} \sigma_{j_2,j_2} = \Theta(1)$ by Condition 3.2.1.

$$\sum_{\substack{(j_1,j_2),\\(j_3,j_4)\in J_A^c}} \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^a \times \mathbf{1}_{\{|\{j_1,j_2\}\cap\{j_3,j_4\}|=2\}} = 2 \sum_{(j_1,j_2)\in J_A^c} (\kappa_1 \sigma_{j_1,j_1}\sigma_{j_2,j_2})^a.$$

Second, if $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 1$, we assume without loss of generality $j_1 = j_3$ and $j_2 \neq j_4$, (B.130) $= \kappa_1 \sigma_{j_1,j_1} \sigma_{j_2,j_4}$, which is nonzero only when $(j_2, j_4) \in J_A$, and then (B.130) $= O(\rho^a)$. By the symmetricity of the indexes,

$$\sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^a \mathbf{1}_{\{|\{j_1,j_2\}\cap\{j_3,j_4\}|=1\}} \leq C \sum_{1\leq j\leq p;\,(j_2,j_4)\in J_A} \rho^a = O(p|J_A|\rho^a).$$

Third, if $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 0$, we know $j_1 \neq j_2 \neq j_3 \neq j_4$, and (B.130) $\neq 0$ only if $(j_1, j_3), (j_2, j_4) \in J_A$ or $(j_1, j_4), (j_2, j_3) \in J_A$. Then (B.130) $= O(\rho^{2a})$. By the symmetricity of the indexes,

$$\sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^a \mathbf{1}_{\{|\{j_1,j_2\}\cap\{j_3,j_4\}|=0\}} \leq C \sum_{(j_1,j_3),(j_2,j_4)\in J_A} \rho^{2a} = O(|J_A|^2\rho^{2a}).$$

In summary, we know

$$\mathrm{var}(T_{U,a,1,1}) = \frac{2a!\kappa_1^a}{P_a^n} \sum_{(j_1,j_2)\in J_A^c} \sigma_{j_1,j_1}^a \sigma_{j_2,j_2}^a + O(p|J_A|\rho^a n^{-a}) + O(|J_A|^2 \rho^{2a} n^{-a}).$$

Since we assume $|J_A|\rho^a = O(pn^{-a/2})$, $|J_A| = o(p^2)$ and $|J_A^c| = \Theta(p^2)$, $\mathrm{var}(T_{U,a,1,1}) \simeq 2a!\kappa_1^a (P_a^n)^{-1} \sum_{1\leq j_1\neq j_2\leq p} (\sigma_{j_1,j_1}\sigma_{j_2,j_2})^a$, which is of order $\Theta(p^2 n^{-a})$.

B.5.12.2 Proof of $\mathrm{var}(T_{U,a,1,2}) = o(p^2 n^{-a})$

As $T_{U,a,1,2} = \sum_{(j_1,j_2)\in J_A^c} \sum_{c=1}^{a} K(c, j_1, j_2)$, by the Cauchy-Schwarz inequality,

$$\mathrm{var}(T_{U,a,1,2}) \leq C \times \sum_{c=1}^{a} \mathrm{var}\Big\{ \sum_{(j_1,j_2)\in J_A^c} K(c, j_1, j_2)\Big\},$$

where C is some constant. As a is finite, to prove $\mathrm{var}(T_{U,a,1,2}) = o(p^2 n^{-a})$, it suffices to prove $\mathrm{var}\{\sum_{(j_1,j_2)\in J_A^c} K(c, j_1, j_2)\} = o(p^2 n^{-a})$, for each $1 \leq c \leq a$. Note that $\mathrm{E}\{K(c, j_1, j_2)\} = 0$ and then

$$\begin{aligned}
\mathrm{var}\Bigg\{ \sum_{(j_1,j_2)\in J_A^c} K(c, j_1, j_2)\Bigg\} &= \mathrm{E}\Big[\Big\{ \sum_{(j_1,j_2)\in J_A^c} K(c, j_1, j_2)\Big\}^2\Big] \\
&= F^2(a, c) \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c);\\ (j_1,j_2),(j_3,j_4)\in J_A^c}} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4),
\end{aligned}$$

where we define

$$\begin{aligned}
Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}\Big[&\prod_{t=1}^{a-c} x_{i_t,j_1} x_{i_t,j_2} \prod_{t=a-c+1}^{a} x_{i_t,j_1} \prod_{t=a+1}^{a+c} x_{i_t,j_2} \\
&\times \prod_{\tilde{t}=1}^{a-c} x_{\tilde{i}_{\tilde{t}},j_3} x_{\tilde{i}_{\tilde{t}},j_4} \prod_{\tilde{t}=a-c+1}^{a} x_{\tilde{i}_{\tilde{t}},j_3} \prod_{\tilde{t}=a+1}^{a+c_2} x_{\tilde{i}_{\tilde{t}},j_4}\Big].
\end{aligned}$$

As $F^2(a, c) = O(n^{-2(a+c)})$, to finish the proof, it remains to prove

$$\sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c);\\ (j_1,j_2),(j_3,j_4)\in J_A^c}} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = o(n^{2(a+c)-a} p^2). \tag{B.131}$$

We note that $\mathrm{E}(x_{1,j}) = 0$ and $\mathrm{E}(x_{1,j_1} x_{1,j_2}) = \mathrm{E}(x_{1,j_3} x_{1,j_4}) = 0$ for $(j_1, j_2), (j_3, j_4) \in J_A^c$. Similarly to Section B.5.12.1, $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = 0$ if $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$, and

$$\sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)} \mathbf{1}_{\{Q_c(\mathbf{i},j_1,j_2,\tilde{\mathbf{i}},j_3,j_4)\neq 0\}} = \sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c)} \mathbf{1}_{\{\{\mathbf{i}\}=\{\tilde{\mathbf{i}}\}\}} = O(n^{a+c}). \tag{B.132}$$

To prove (B.131), it remains to prove for given $\mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)$,

$$\Big| \sum_{(j_1,j_2),(j_3,j_4) \in J_A^c} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \Big| = O(p^2). \tag{B.133}$$

We next prove (B.133) by discussing the value of $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)$. To facilitate the discussion, for given $\mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)$, we decompose the sets $\{\mathbf{i}\}$ and $\{\tilde{\mathbf{i}}\}$ into three disjoint sets respectively, defined as

$$\{\mathbf{i}\}_{(1)} = \{i_1, \ldots, i_{a-c}\},\ \{\mathbf{i}\}_{(2)} = \{i_{a-c+1}, \ldots, i_a\},\ \{\mathbf{i}\}_{(3)} = \{i_{a+1}, \ldots, i_{a+c}\},$$
$$\{\tilde{\mathbf{i}}\}_{(1)} = \{\tilde{i}_1, \ldots, \tilde{i}_{a-c}\},\ \{\tilde{\mathbf{i}}\}_{(2)} = \{\tilde{i}_{a-c+1}, \ldots, \tilde{i}_a\},\ \{\tilde{\mathbf{i}}\}_{(3)} = \{\tilde{i}_{a+1}, \ldots, \tilde{i}_{a+c}\},$$

which satisfy that $\{\mathbf{i}\} = \cup_{l=1}^3 \{\mathbf{i}\}_{(l)}$ and $\{\tilde{\mathbf{i}}\} = \cup_{l=1}^3 \{\tilde{\mathbf{i}}\}_{(l)}$.

When $c \leq a-1$, $\{\mathbf{i}\}_{(1)} \neq \emptyset$. We consider an index $i \in \{\mathbf{i}\}_{(1)}$, and discuss four different cases. First, if $i \notin \{\tilde{\mathbf{i}}\}$,

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1} x_{i,j_2}) \mathrm{E}(\text{other terms}) = 0,$$

where the last equation follows from $\mathrm{E}(x_{i,j_1} x_{i,j_2}) = 0$ when $(j_1, j_2) \in J_A$. Second, if $i \in \{\tilde{\mathbf{i}}\}_{(2)}$,

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1} x_{i,j_2} x_{i,j_3}) \mathrm{E}(\text{other terms}) = 0,$$

where the last equation is obtained by Condition 3.2.6. Third, if $i \in \{\tilde{\mathbf{i}}\}_{(3)}$, similarly by Condition 3.2.6, we also know

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1} x_{i,j_2} x_{i,j_4}) \mathrm{E}(\text{other terms}) = 0. \tag{B.134}$$

Fourth, if $i \in \{\tilde{\mathbf{i}}\}_{(1)}$,

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1}x_{i,j_2}x_{i,j_3}x_{i,j_4})\mathrm{E}(\text{other terms}). \tag{B.135}$$

Under Condition 3.2.6, as $\mathrm{E}(x_{i,j_1}x_{i,j_2}) = \mathrm{E}(x_{i,j_3}x_{i,j_4}) = 0$ when (j_1, j_2) and $(j_3, j_4) \in J_A^c$,

$$\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big) = \kappa_1\Big\{\mathrm{E}(x_{i,j_1}x_{i,j_3})\mathrm{E}(x_{i,j_2}x_{i,j_4}) + \mathrm{E}(x_{i,j_1}x_{i,j_4})\mathrm{E}(x_{i,j_2}x_{i,j_3})\Big\}.$$

In addition, when $c = a$, $\{\mathbf{i}\}_{(1)} = \emptyset$ but $\{\mathbf{i}\}_{(2)}$ and $\{\mathbf{i}\}_{(3)} \neq \emptyset$. We next consider an index $i \in \{\mathbf{i}\}_{(2)}$ without loss of generality. Following similar analysis, we know $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = 0$ when $i \notin \{\tilde{\mathbf{i}}\}$.

By symmetrically analyzing the indexes in $\mathbf{i}$ and $\tilde{\mathbf{i}}$ similarly as above, we know that $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \neq 0$ only when $\{\mathbf{i}\}_{(1)} = \{\tilde{\mathbf{i}}\}_{(1)}$ and $\{\mathbf{i}\}_{(2)} \cup \{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$. When $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \neq 0$, suppose $r = |\{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(2)}|$ then $|\{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(3)}| = c - r$, $|\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(2)}| = c - r$, and $|\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(3)}| = r$. It follows that

$$\begin{aligned} &Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \qquad\qquad (\text{B.136})\\ =&\Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^{a-c}\{\mathrm{E}(x_{1,j_1}x_{1,j_3})\mathrm{E}(x_{1,j_2}x_{1,j_4})\}^r\{\mathrm{E}(x_{1,j_1}x_{1,j_4})\mathrm{E}(x_{1,j_2}x_{1,j_3})\}^{c-r}\\ =&\Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{1,j_t}\Big)\Big\}^{a-c}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^r(\sigma_{j_1,j_4}\sigma_{j_2,j_3})^{c-r}. \end{aligned}$$

To prove (B.133), we next examine the value of (B.136) with respect to three different cases of (j_1, j_2, j_3, j_4).

Case (1) If $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 2$, it means that $\{j_1, j_2\} = \{j_3, j_4\}$. Assume, without loss of generality, that $j_1 = j_3$ and $j_2 = j_4$. Then (B.136) $= O\{(\sigma_{j_1,j_1}\sigma_{j_2,j_2})^{a-c+r} \times (\sigma_{j_1,j_2}^2)^{c-r}\}$, which is nonzero only when $r = c$ as $\sigma_{j_1,j_2} = 0$. By the symmetricity of j

indexes and the boundedness of moments in Condition 3.2.1,

$$\left| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{|\{j_1,j_2\}\cap\{j_3,j_4\}|=2\}} \right| \leq Cp^2.$$

Case (2) If $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 1$, we assume without loss of generality that $j_1 = j_3$ but $j_2 \neq j_4$. Then (B.136) $= O(1)(\sigma_{j_1,j_1}\sigma_{j_2,j_4})^{a-c+r}(\sigma_{j_1,j_4}\sigma_{j_1,j_2})^{c-r}$, which is also nonzero only when $r = c$. By the symmetricity of j indexes and Condition 3.2.1, we have

$$\begin{aligned} & \left| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{|\{j_1,j_2\}\cap\{j_3,j_4\}|=1\}} \right| \\ \leq \; & C\Big| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} (\sigma_{j_1,j_1}\sigma_{j_2,j_4})^a \Big| \leq \sum_{1\leq j\leq p,\,(j_2,j_4)\in J_A} O(\rho^a) = O(p|J_A|\rho^a), \end{aligned}$$

where we use Condition 3.2.5 that $\sigma_{j_2,j_4} = \rho$ when $(j_2, j_4) \in J_A$ and $\sigma_{j_2,j_4} = 0$ when $(j_2, j_4) \notin J_A$.

Case (3) If $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 0$, it means that $j_1 \neq j_2 \neq j_3 \neq j_4$. Then

$$\text{(B.136)} = O(1)(\sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3})^{a-c}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^r(\sigma_{j_1,j_4}\sigma_{j_2,j_3})^{c-r},$$

which nonzero only when $(j_1, j_3), (j_2, j_4) \in J_A^c$ or $(j_1, j_4), (j_2, j_3) \in J_A^c$. By the symmetricity of j indexes, Condition 3.2.1 and Condition 3.2.5,

$$\begin{aligned} & \left| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{|\{j_1,j_2\}\cap\{j_3,j_4\}|=0\}} \right| \\ \leq \; & C \sum_{(j_1,j_3),(j_2,j_4)\in J_A^c} \rho^{2a} = O(|J_A|^2\rho^{2a}). \end{aligned}$$

In summary,

$$\Big| \sum_{(j_1,j_2),(j_3,j_4)\in J_A^c} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \Big| = O(p^2 + p|J_A|\rho^a + |J_A|^2\rho^{2a}) = o(p^2),$$

as we assume $|J_A|\rho^a = O(pn^{-a/2})$.

B.5.12.3 Proof of $\mathrm{var}(T_{U,a,2}) = o(p^2 n^{-a})$

Similarly to Section B.5.12.2, by the Cauchy-Schwarz inequality,

$$\mathrm{var}(T_{U,a,2}) \leq C \sum_{c=0}^{a} \mathrm{var}(T_{U,a,2,c}), \tag{B.137}$$

where $T_{U,a,2,c} = \sum_{(j_1,j_2)\in J_A} K(c, j_1, j_2)$. To prove $\mathrm{var}(T_{U,a,2}) = o(p^2 n^{-a})$, it suffices to prove $\mathrm{var}(T_{U,a,2,c}) = o(p^2 n^{-a})$ for $0 \leq c \leq a$. Following the notation in Section B.5.12.2, we have

$$\mathrm{E}(T_{U,a,2,c}^2) = F^2(a,c) \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c);\\ (j_1,j_2),(j_3,j_4)\in J_A}} Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4).$$

When $1 \leq c \leq a$, $\mathrm{E}(T_{U,a,2,c}) = 0$; when $c = 0$, $\mathrm{E}(T_{U,a,2,0}) = \sum_{(j_1,j_2)\in J_A} \sigma_{j_1,j_2}^a$. Then

$$\mathrm{var}(T_{U,a,2,c}) = F^2(a,c) \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a+c);\\ (j_1,j_2),(j_3,j_4)\in J_A}} \tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4), \tag{B.138}$$

where we define $\tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)$ when $1 \leq c \leq a$; and $\tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) - (\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a$ when $c = 0$.

To prove $\mathrm{var}(T_{U,a,2,c}) = o(p^2 n^{-a})$ for $1 \leq c \leq a$, we next examine the value of $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)$. For given $\mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n, a+c)$, we define $\{\mathbf{i}\}_{(l)}$ and $\{\tilde{\mathbf{i}}\}_{(l)}$ for

$l = 1, 2, 3$ same as in Section B.5.12.2. Consider an index $i \in \{\mathbf{i}\}_{(2)}$. If $i \notin \{\tilde{\mathbf{i}}\}$,

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1})\mathrm{E}(\text{other terms}) = 0.$$

If $i \in \{\tilde{\mathbf{i}}\}_{(1)}$, by Condition 3.2.6,

$$Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = \mathrm{E}(x_{i,j_1}x_{i,j_3}x_{i,j_4})\mathrm{E}(\text{other terms}) = 0.$$

Similarly, for an index $i \in \{\mathbf{i}\}_{(3)}$, we have $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = 0$ if $i \notin \{\tilde{\mathbf{i}}\}$ or $i \in \{\tilde{\mathbf{i}}\}_{(1)}$. Analyzing the indexes in $\{\tilde{\mathbf{i}}\}$ symmetrically, we know that $Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \neq 0$ only when $\{\mathbf{i}\}_{(2)} \cup \{\mathbf{i}\}_{(3)} = \{\tilde{\mathbf{i}}\}_{(2)} \cup \{\tilde{\mathbf{i}}\}_{(3)}$. Suppose $|\{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(2)}| = r$, then $|\{\mathbf{i}\}_{(2)} \cap \{\tilde{\mathbf{i}}\}_{(3)}| = c - r$, $|\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(2)}| = c - r$, and $|\{\mathbf{i}\}_{(3)} \cap \{\tilde{\mathbf{i}}\}_{(3)}| = r$. Moreover, we let $|\{\mathbf{i}\}_{(1)} \cap \{\tilde{\mathbf{i}}\}_{(1)}| = t_c$ then $0 \leq t_c \leq a - c$. It follows that

$$\begin{aligned}
& Q_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \qquad \text{(B.139)} \\
= \; & \Big\{\mathrm{E}\Big(\prod_{t=1}^{4} x_{i,j_t}\Big)\Big\}^{t_c} \Big\{\mathrm{E}(x_{i,j_1}x_{i,j_2})\mathrm{E}(x_{i,j_3}x_{i,j_4})\Big\}^{a-c-t_c} \\
& \times\{\mathrm{E}(x_{i,j_1}x_{i,j_3})\mathrm{E}(x_{i,j_2}x_{i,j_4})\}^r\{\mathrm{E}(x_{i,j_1}x_{i,j_4})\mathrm{E}(x_{i,j_2}x_{i,j_3})\}^{c-r}.
\end{aligned}$$

To examine (B.131), we next analyze (B.139) with respect to different c and t_c values, where $0 \leq c \leq a$, $0 \leq r \leq c$, and $0 \leq t_c \leq a - c$.

When $c = 0$ and $t_c = t_0 = 0$, it means that $\{\mathbf{i}\} = \{\mathbf{i}\}_{(1)}$, $\{\tilde{\mathbf{i}}\} = \{\tilde{\mathbf{i}}\}_{(1)}$, $\{\mathbf{i}\} \cap \{\tilde{\mathbf{i}}\} = \emptyset$, and $Q_0(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) = (\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a$. Then

$$\begin{aligned}
& \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \tilde{Q}_0(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \times \mathbf{1}_{\{t_0=0\}} \qquad \text{(B.140)} \\
= & \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \Big\{Q_0(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) - (\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a\Big\}\mathbf{1}_{\{t_0=0\}} = 0.
\end{aligned}$$

In the following, it remains to consider the cases when $c \geq 1$ or $t_c \geq 1$ in (B.139), which are examined by discussing three cases (j_1, j_2, j_3, j_4) below.

Case (1) If $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 2$, we assume without loss of generality that $j_1 = j_3$ and $j_2 = j_4$. Then by Condition 3.2.6, $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) = \kappa_1(2\sigma_{j_1,j_2}^2 + \sigma_{j_1,j_1}\sigma_{j_2,j_2})$, and (B.139) $= \{\kappa_1(2\sigma_{j_1,j_2}^2 + \sigma_{j_1,j_1}\sigma_{j_2,j_2})\}^{t_c}\sigma_{j_1,j_2}^{2(a-c-t_c)}(\sigma_{j_1,j_1}\sigma_{j_2,j_2})^r(\sigma_{j_1,j_2})^{2(c-r)}$.

Case (1.1) For $c = 0$ and $1 \leq t_c = t_0 \leq a$, we have $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \leq 2a - t_0$, and

$$\begin{aligned}
&\left| \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a);\\ (j_1,j_2),(j_3,j_4)\in J_A}} \tilde{Q}_0(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{c=0,1\leq t_0\leq a,|\{j_1,j_2\}\cap\{j_3,j_4\}|=2\}} \right| \qquad \text{(B.141)}\\
\leq\ & C\sum_{t_0=1}^{a} n^{2a-t_0} \sum_{(j_1,j_2)\in J_A} |\sigma_{j_1,j_2}|^{2(a-t_0)}|2\sigma_{j_1,j_2}^2 + \sigma_{j_1,j_1}\sigma_{j_2,j_2}|^{t_0} + |\sigma_{j_1,j_2}|^{2a}\\
=\ & \sum_{t_0=1}^{a} O(1)n^{2a-t_0}|J_A| \times (\rho^{2a} + \rho^{2(a-t_0)}),
\end{aligned}$$

where we use Condition 3.2.5.

Case (1.2) For $1 \leq c \leq a$ and $0 \leq t_c \leq a - c$, we have $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \leq 2a - t_c$, and for each c given,

$$\begin{aligned}
&\left| \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n,a);\\ (j_1,j_2),(j_3,j_4)\in J_A}} \tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4)\mathbf{1}_{\{1\leq t_c\leq a-c,|\{j_1,j_2\}\cap\{j_3,j_4\}|=2\}} \right| \qquad \text{(B.142)}\\
\leq\ & C \sum_{\substack{0\leq r\leq c;\\ 0\leq t_c\leq a-c}} n^{2a-t_c} \sum_{(j_1,j_2)\in J_A} |\sigma_{j_1,j_2}|^{2(a-c-t_c)}\\
& \times|2\sigma_{j_1,j_2}^2 + \sigma_{j_1,j_1}\sigma_{j_2,j_2}|^{t_c}|\sigma_{j_1,j_1}\sigma_{j_2,j_2}|^r|\sigma_{j_1,j_2}|^{2(c-r)}\\
=\ & \sum_{\substack{0\leq r\leq c;\\ 0\leq t_c\leq a-c}} O(1)n^{2a-t_c}|J_A|\{\rho^{2(a-r)} + \rho^{2(a-t_c-r)}\}.
\end{aligned}$$

Case (2) If $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 1$, we assume without loss of generality that $j_1 = j_3$ and $j_2 \neq j_4$. Then by Condition 3.2.6, $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) = \kappa_1(2\sigma_{j_1,j_2}\sigma_{j_1,j_4} +$

$\sigma_{j_1,j_1}\sigma_{j_2,j_4}$). We then know

$$\text{(B.139)} = \{\kappa_1(2\sigma_{j_1,j_2}\sigma_{j_1,j_4} + \sigma_{j_1,j_1}\sigma_{j_2,j_4})\}^{t_c}(\sigma_{j_1,j_2}\sigma_{j_1,j_4})^{a-c-t_c}(\sigma_{j_1,j_1}\sigma_{j_2,j_4})^r(\sigma_{j_1,j_4}\sigma_{j_1,j_2})^{c-r}.$$

Case (2.1) For $c = 0$ and $1 \leq t_c = t_0 \leq a$, we have $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \leq 2a - t_0$, and (B.139) $\neq 0$ at least when $(j_1, j_2), (j_1, j_4) \in J_A$. Then

$$\begin{aligned} & \Bigg| \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \tilde{Q}_0(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{c=0, 1 \leq t_0 \leq a, |\{j_1,j_2\} \cap \{j_3,j_4\}|=1\}} \Bigg| \qquad \text{(B.143)} \\ \leq \; & C \sum_{t_0=1}^{a} n^{2a-t_0} \sum_{(j_1,j_2),(j_1,j_4) \in J_A} \Big(|\sigma_{j_1,j_2}\sigma_{j_1,j_4}|^a + |\sigma_{j_2,j_4}|^{t_0} |\sigma_{j_1,j_2}\sigma_{j_1,j_4}|^{a-t_0} \Big) \\ = \; & \sum_{t_0=1}^{a} O(1) n^{2a-t_0} \max_{1 \leq j_1 \leq p} |J_{j_1}| \times |J_A| (\rho^{2a} + \rho^{2a-t_0}). \end{aligned}$$

Case (2.2) For $c \geq 1$ and $0 \leq t_c \leq a - c$, we have $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \leq 2a - t_c$. (B.139) $\neq 0$ when $(j_1, j_2), (j_1, j_4) \in J_A$ or $(j_2, j_4) \in J_A$. For given c, the range of (B.139) is between $O(\rho^{2a-t_c-r})$ and $O(\rho^{2a-r})$.

$$\begin{aligned} & \Bigg| \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{0 \leq t_c \leq a-c, |\{j_1,j_2\} \cap \{j_3,j_4\}|=1\}} \Bigg| \qquad \text{(B.144)} \\ = \; & \sum_{\substack{0 \leq r \leq c; \\ 0 \leq t_c \leq a-c}} O(1) n^{2a-t_c} \max_{1 \leq j_1 \leq p} |J_{j_1}| \times |J_A| (\rho^{2a-t_c-r} + \rho^{2a-r}). \end{aligned}$$

Case (3) If $|\{j_1, j_2\} \cap \{j_3, j_4\}| = 0$, we know $j_1 \neq j_2 \neq j_3 \neq j_4$. Then by Condition 3.2.6 and 3.2.5, $\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) = \kappa_1(\sigma_{j_1,j_2}\sigma_{j_3,j_4} + \sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3}) = O(\rho^2)$. Therefore, (B.139) $= O(\rho^{2a})$.

Case (3.1) For $c = 0$ and $1 \leq t_c = t_0 \leq a$, we have $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \leq 2a - t_0$.

$$\left| \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \tilde{Q}_0(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{c=0, 1\leq t_0 \leq a, |\{j_1,j_2\} \cap \{j_3,j_4\}|=0\}} \right| \tag{B.145}$$
$$\leq \quad C \sum_{t_0=1}^{a} \sum_{(j_1,j_2),(j_3,j_4)\in J_A} |\sigma_{j_1,j_2}\sigma_{j_3,j_4}|^a = \sum_{t_0=1}^{a} n^{2a-t_0} |J_A|^2 O(\rho^{2a}).$$

Case (3.2) For $1 \leq c \leq a$ and $0 \leq t_c \leq a$, we have $|\{\mathbf{i}\} \cup \{\tilde{\mathbf{i}}\}| \leq 2a - t_c$. Then for given $c \geq 1$,

$$\left| \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}} \in \mathcal{P}(n,a); \\ (j_1,j_2),(j_3,j_4) \in J_A}} \tilde{Q}_c(\mathbf{i}, j_1, j_2, \tilde{\mathbf{i}}, j_3, j_4) \mathbf{1}_{\{0\leq t_c \leq a-c, |\{j_1,j_2\} \cap \{j_3,j_4\}|=0\}} \right| \tag{B.146}$$
$$\leq \quad C \sum_{\substack{0\leq r\leq c; \\ 0\leq t_c \leq a-c}} n^{2a-t_c} \sum_{(j_1,j_2),(j_3,j_4)\in J_A} |\sigma_{j_1,j_2}\sigma_{j_3,j_4}|^a = \sum_{t_c=0}^{a-c} n^{2a-t_c} |J_A|^2 O(\rho^{2a}),$$

where we use the symmetricity of indexes.

Combining (B.141)–(B.146) above, and by (B.137) and (B.138) and $F(a,c) = O(n^{-(a+c)})$, we know

$$\begin{aligned}
\mathrm{var}(T_{1,a,2}) \quad = \quad & \sum_{t_0=1}^{a} O(1) \frac{1}{n^{t_0}} |J_A| \times \{\rho^{2a} + \rho^{2(a-t_0)}\} \qquad \text{(B.147)} \\
& + \sum_{c=1}^{a} \sum_{t_c=0}^{a-c} \sum_{r=0}^{c} O(1) |J_A| \frac{1}{n^{2c+t_c}} \{\rho^{2(a-r)} + \rho^{2(a-t_c-r)}\} \\
& + \sum_{t_0=1}^{a} O(1) \frac{1}{n^{t_0}} \max_{1\leq j_1 \leq p} |J_{j_1}| \times |J_A| (\rho^{2a} + \rho^{2a-t_0}) \\
& + \sum_{c=1}^{a} \sum_{t_c=0}^{a-c} \sum_{r=0}^{c} O(1) \frac{1}{n^{2c+t_c}} \max_{1\leq j_1 \leq p} |J_{j_1}| |J_A| (\rho^{2a-t_c-r} + \rho^{2a-r}) \\
& + \sum_{t_0=1}^{a} O(1) \frac{1}{n^{t_0}} |J_A|^2 \rho^{2a} + \sum_{c=1}^{a} \sum_{t_c=0}^{a-c} O(1) \frac{1}{n^{2c+t_c}} |J_A|^2 \rho^{2a}.
\end{aligned}$$

We then examine the six summed terms in the right hand side of (B.147) and show that they are $o(p^2 n^{-a})$ respectively.

(1) For the first term in (B.147), as $|J_A|\rho^a = O(pn^{-a/2})$, $n^{-t_0}|J_A|\rho^{2a} = n^{-t_0} \times |J_A|^{-1}|J_A|^2\rho^{2a} = o(p^2 n^{-a})$, and

$$\begin{aligned} n^{-t_0}|J_A|\rho^{2(a-t_0)} =& n^{-t_0}|J_A|^{1-2(a-t_0)/a}(|J_A|\rho^a)^{2(a-t_0)/a} \\ =& O(1)p^2 n^{-a}|J_A|^{-1+t_0/a}(|J_A|/p^2)^{t_0/a} = o(p^2 n^{-a}), \end{aligned}$$

where we use $1 \le t_0 \le a$ and $|J_A| = o(p^2)$ in the last equation.

(2) For the second term in (B.147), as $r \le c \le a$ and $|J_A| = o(p^2)$,

$$\begin{aligned} n^{-(2c+t_c)}|J_A|\rho^{2(a-r)} =& n^{-(2c+t_c)}|J_A|^{1-2(a-r)/a}(|J_A|\rho^a)^{2(a-r)/a} \\ =& O(1)p^2 n^{-a+r-2c-t_c}|J_A|^{-1+r/a}(|J_A|/p^2)^{r/a} = o(p^2 n^{-a}), \end{aligned}$$

and similarly as $r \le c \le a$, $t_c + r \le a$ and $c \ge 1$,

$$\begin{aligned} & n^{-(2c+t_c)}|J_A|\rho^{2(a-t_c-r)} \\ =& O(1)p^2 n^{-a+t_c+r-2c-t_c}|J_A|^{-1+(t_c+r)/a}(|J_A|/p^2)^{(t_c+r)/a} = o(p^2 n^{-a}). \end{aligned}$$

(3) For the third term in (B.147), as $1 \le t_0 \le a$, and $|J_A|\rho^a = O(pn^{-a/2})$,

$$n^{-t_0} \max_{1 \le j_1 \le p} |J_{j_1}| \times |J_A|\rho^{2a} = \frac{\max_{1 \le j_1 \le p} |J_{j_1}|}{n^{t_0}|J_A|}|J_A|^2\rho^{2a} = o(p^2 n^{-a}),$$

and

$$
\begin{aligned}
& n^{-t_0} \max_{1 \le j_1 \le p} |J_{j_1}| \times |J_A| \rho^{2a-t_0} \\
= & n^{-t_0} \max_{1 \le j_1 \le p} |J_{j_1}| \times |J_A|^{1-(2a-t_0)/a} (|J_A|\rho^a)^{(2a-t_0)/a} \\
= & O(1) p^2 n^{-a-t_0/2} \max_{1 \le j_1 \le p} |J_{j_1}| \times |J_A|^{-1+t_0/a} p^{-t_0/a} \\
= & O(1) \frac{p^2}{n^{a+t_0/2}} \Big(\frac{\max_{1 \le j_1 \le p} |J_{j_1}|}{|J_A|} \Big)^{1-t_0/a} \Big(\frac{\max_{1 \le j_1 \le p} |J_{j_1}|}{p} \Big)^{t_0/a} = o(p^2 n^{-a}),
\end{aligned}
$$

where in the last equation, we use the facts that $1 \le t_0 \le a$, $\max_{1 \le j_1 \le p} |J_{j_1}| \le |J_A|$, and $\max_{1 \le j_1 \le p} |J_{j_1}| \le p$.

(4) For the fourth term in (B.147),

$$
\begin{aligned}
& n^{-(2c+t_c)} \max_{1 \le j_1 \le p} |J_{j_1}| |J_A| \rho^{2a-t_c-r} \\
= & n^{-(2c+t_c)} \max_{1 \le j_1 \le p} |J_{j_1}| |J_A|^{1-(2a-t_c-r)/a} (|J_A|\rho^a)^{(2a-t_c-r)/a} \\
= & O(1) \frac{p^2}{n^a} \frac{1}{n^{2c+t_c/2-r/2}} \frac{\max_{1 \le j_1 \le p} |J_{j_1}|}{|J_A|} \Big(\frac{|J_A|}{p} \Big)^{(t_c+r)/a} \\
= & O(1) \frac{p^2}{n^a} \frac{1}{n^{2c+t_c/2-r/2}} \Big(\frac{\max_{1 \le j_1 \le p} |J_{j_1}|}{|J_A|} \Big)^{1-(t_c+r)/a} \Big(\frac{\max_{1 \le j_1 \le p} |J_{j_1}|}{p} \Big)^{(t_c+r)/a} = o(p^2 n^{-a}),
\end{aligned}
$$

where we obtain the last equation by noting that $t_c + r \le a$, $r \le c$, and $c \ge 1$. Similarly, we have

$$
\begin{aligned}
& n^{-(2c+t_c)} \max_{1 \le j_1 \le p} |J_{j_1}| |J_A| \rho^{2a-r} \\
= & O(1) \frac{p^2}{n^a} \frac{1}{n^{2c+t_c-r/2}} \Big(\frac{\max_{1 \le j_1 \le p} |J_{j_1}|}{|J_A|} \Big)^{1-r/a} \Big(\frac{\max_{1 \le j_1 \le p} |J_{j_1}|}{p} \Big)^{r/a} = o(p^2 n^{-a}).
\end{aligned}
$$

(5) For the fifth and sixth terms in (B.147), as $|J_A|\rho^a = O(pn^{-a/2})$, $t_0 \ge 1$ and $c \ge 1$, we know $n^{-t_0}|J_A|^2\rho^{2a} = o(p^2 n^{-a})$, and $n^{-(2c+t_c)}|J_A|^2\rho^{2a} = o(p^2 n^{-a})$.

B.5.13 Proof of Lemma B.1.12

The proof is similar to Section B.5.3. In particular, Lemma B.1.12 shows that $\text{var}\{\mathcal{U}(a)\} \simeq \text{var}(T_{U,a,1,1})$. By the Cauchy-schwarz inequality,

$$\text{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} = \text{E}\{T_{U,a,1,1}T_{U,b,1,1}\}/\{\sigma(a)\sigma(b)\} + o(1),$$

where we use $\text{E}(T_{U,a,1,1}) = \text{E}(T_{U,b,1,1}) = 0$. For two integers $a \neq b$, we next prove $\text{E}(T_{U,a,1,1}T_{U,b,1,1})$=0. Specifically,

$$\text{E}(T_{U,a,1,1}T_{U,b,1,1}) = (P_a^n P_b^n)^{-1} \sum_{\substack{\mathbf{i}\in\mathcal{P}(n,a),\tilde{\mathbf{i}}\in\mathcal{P}(n,b); \\ (j_1,j_2),(j_3,j_4)\in J_A^c}} \text{E}\Big(\prod_{k=1}^{a} x_{i_k,j_1}x_{i_k,j_2}\prod_{\tilde{k}=1}^{b} x_{\tilde{i}_{\tilde{k}},j_3}x_{\tilde{i}_{\tilde{k}},j_4}\Big).$$

Since $a \neq b$, $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$. Assume without loss of generality that $a < b$ and index $i \in \{\tilde{\mathbf{i}}\}$ but $i \notin \{\mathbf{i}\}$. Then

$$\text{E}\Big(\prod_{k=1}^{a} x_{i_k,j_1}x_{i_k,j_2}\prod_{\tilde{k}=1}^{b} x_{\tilde{i}_{\tilde{k}},j_3}x_{\tilde{i}_{\tilde{k}},j_4}\Big) = \text{E}(x_{1,j_3}x_{1,j_4}) \times \text{E}(\text{other terms}) = 0,$$

where we use the $\sigma_{j_1,j_2} = \sigma_{j_3,j_4} = 0$ for $(j_1,j_2),(j_3,j_4) \in J_A^c$. Therefore, we have $\text{cov}(T_{U,a,1,1}, T_{U,b,1,1}) = 0$ and the lemma is proved.

B.5.14 Proof of Lemma B.1.13

We prove Lemma B.1.13 similarly as in Section B.5.6. By the Cauchy-Schwarz inequality, for some constant C, $\text{var}(\sum_{k=1}^n \pi_{n,k}^2) \leq Cn^2 \max_{1\leq k\leq n;\, 1\leq r_1,r_2\leq m} \text{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}})$, where $c(n,a) = [a \times \{\sigma(a)P_a^n\}^{-1}]^2$ and for two finite integers a_1 and a_2, $\mathbb{T}_{k,a_1,a_2} = \text{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})$. In particular, when $k < \max\{a_1,a_2\}$, $\mathbb{T}_{k,a_1,a_2} = 0$; when $k \geq$

$\max\{a_1, a_2\}$,

$$\begin{aligned}\mathbb{T}_{k,a_1,a_2} &= \mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})\\ &= \sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\,l=1,2;\\(j_1,j_2),(j_3,j_4)\in J_A^c}} \Big\{\prod_{l=1}^{2} c(n,a_l)\Big\}^{1/2} \mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\end{aligned}$$

with $\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l = 1,2) = \mathrm{E}(\prod_{t=1}^{4} x_{k,j_t})\prod_{l=1}^{2}\prod_{t=1}^{a_l-1}(x_{i_t^{(l)},j_{2l-1}}x_{i_t^{(l)},j_{2l}})$. To prove $\mathrm{var}(\sum_{k=1}^{n}\pi_{n,k}^2) \to 0$, it suffices to prove $\mathrm{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}}) = o(n^{-2})$ for any $1 \le r_1, r_2 \le m$. Without loss of generality, we consider two finite integers a_1 and a_2, and prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$ when $\max\{a_1, a_2\} \le k \le n$.

To prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$, we decompose $\mathbb{T}_{k,a_1,a_2} = \sum_{M=2}^{4}\mathbb{T}_{k,a_1,a_2,(M)}$, where

$$\begin{aligned}\mathbb{T}_{k,a_1,a_2,(M)} = &\sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\,l=1,2;\\(j_1,j_2),(j_3,j_4)\in J_A^c}} \mathbf{1}_{\{|\{j_1,j_2\}\cup\{j_3,j_4\}|=M\}}\\ &\times\Big\{\prod_{l=1}^{2} c(n,a_l)\Big\}^{1/2}\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2).\end{aligned}$$

Here $2 \le M \le 4$ because $2 \le |\{j_1, j_2\} \cup \{j_3, j_4\}| \le 4$ when $(j_1, j_2), (j_3, j_4) \in J_A^c$. By the Cauchy-Schwarz inequality, to prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$, it suffices to prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,(M)}) = o(n^{-2})$ for $M = 2, 3, 4$. For easy presentation, we let $a_3 = a_1$ and $a_4 = a_2$, and then

$$\begin{aligned}\mathbb{T}^2_{k,a_1,a_2,(M)} = &\sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\,l=1,2,3,4;\\(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{\left\{\substack{|\{j_1,j_2\}\cup\{j_3,j_4\}|=M,\\|\{j_5,j_6\}\cup\{j_7,j_8\}|=M}\right\}}\\ &\times\Big\{\prod_{l=1}^{2} c(n,a_l)\Big\}\times\mathbb{X}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2,3,4),\end{aligned}$$

where

$$
\begin{aligned}
&\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4) \\
&= \mathrm{E}\Big(\prod_{t=1}^{4} x_{k,j_t}\Big)\mathrm{E}\Big(\prod_{t=5}^{8} x_{k,j_t}\Big)\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)}, j_{2l-1}} x_{i_t^{(l)}, j_{2l}}\Big).
\end{aligned}
$$

By $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\} = \mathrm{E}\{\mathbb{T}^2_{k,a_1,a_2,(M)}\} - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,(M)})\}^2$,

$$
\begin{aligned}
&\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\} \\
= &\sum_{\substack{\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l-1),\, l=1,2,3,4; \\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in J_A^c}} \mathbf{1}_{\left\{\substack{|\{j_1,j_2\}\cup\{j_3,j_4\}|=M, \\ |\{j_5,j_6\}\cup\{j_7,j_8\}|=M}\right\}} \Big\{\prod_{l=1}^{2} c(n, a_l)\Big\} \\
&\times \Big[\mathrm{E}\Big\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\Big\} \\
&\quad - \mathrm{E}\Big\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2)\Big\} \times \mathrm{E}\Big\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 3, 4)\Big\}\Big],
\end{aligned}
$$

where we similarly define

$$
\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 3, 4) = \mathrm{E}\Big(\prod_{t=5}^{8} x_{k,j_t}\Big)\prod_{l=3}^{4}\prod_{t=1}^{a_l-1}(x_{i_t^{(l)}, j_{2l-1}} x_{i_t^{(l)}, j_{2l}}).
$$

To prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,(M)}) = o(n^{-2})$, we examine the value of

$$
\begin{aligned}
&\mathrm{E}\Big\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\Big\} \\
&- \mathrm{E}\Big\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2)\Big\}\mathrm{E}\Big\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 3, 4)\Big\}.
\end{aligned} \tag{B.148}
$$

We next show that when (B.148) $\neq 0$, the following two claims hold:

$$
\begin{aligned}
&\textit{Claim 1:}\quad (\{\mathbf{i}^{(1)}\} \cup \{\mathbf{i}^{(2)}\}) \cap (\{\mathbf{i}^{(3)}\} \cup \{\mathbf{i}^{(4)}\}) \neq \emptyset, \\
&\textit{Claim 2:}\ |\cup_{l=1}^{4} \{\mathbf{i}^{(l)}\}| \leq a_1 + a_2 - 2.
\end{aligned} \tag{B.149}
$$

Claim 1 can be straightforwardly seen from the definition (B.148). We then prove

Claim 2. Note that $\mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\} \neq 0$ only when $|\cup_{l=1}^{4} \{\mathbf{i}^{(l)}\}| \leq a_1 + a_2 - 2$ following similar analysis to Section B.5.6.1. In addition, as $\sigma_{j_1,j_2} = \sigma_{j_3,j_4} = 0$ when $(j_1, j_2), (j_3, j_4) \in J_A^c$, we know that $\mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2)\} \neq 0$ only when $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$; as $\sigma_{j_5,j_6} = \sigma_{j_7,j_8} = 0$, we similarly know that $\mathrm{E}\{\mathbb{X}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 3, 4)\} \neq 0$ only when $\{\mathbf{i}^{(3)}\} = \{\mathbf{i}^{(4)}\}$. It follows that if $|\cup_{l=1}^{4} \{\mathbf{i}^{(l)}\}| > a_1 + a_2 - 2$, (B.148) $= 0$. Thus to evaluate $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}$, it remains to consider (B.148) under the cases when $(\{\mathbf{i}^{(1)}\} \cup \{\mathbf{i}^{(2)}\}) \cap (\{\mathbf{i}^{(3)}\} \cup \{\mathbf{i}^{(4)}\}) \neq \emptyset$ and $|\cup_{l=1}^{4} \{\mathbf{i}^{(l)}\}| \leq a_1 + a_2 - 2$.

Given the two claims above, we examine $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}$ for $M = 2, 3, 4$ respectively. In particular, we decompose $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\} = \mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)} + \mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(2)}$, where

$$\begin{aligned}
&\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)} \\
&= \sum_{\substack{\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1),\, l=1,2,3,4; \\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in J_A^c}} \mathbf{1}_{\left\{\substack{|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}| = a_1+a_2-2; \\ |\{j_1,j_2\}\cup\{j_3,j_4\}| = M; \\ |\{j_5,j_6\}\cup\{j_7,j_8\}| = M}\right\}} \prod_{l=1}^{2} c(n, a_l) \times \text{(B.148)},
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(2)} \\
&= \sum_{\substack{\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1),\, l=1,2,3,4; \\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in J_A^c}} \mathbf{1}_{\left\{\substack{|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}| < a_1+a_2-2; \\ |\{j_1,j_2\}\cup\{j_3,j_4\}| = M; \\ |\{j_5,j_6\}\cup\{j_7,j_8\}| = M}\right\}} \prod_{l=1}^{2} c(n, a_l) \times \text{(B.148)}.
\end{aligned}$$

We next consider $M = 2, 3, 4$ in the following **Cases (1)–(3)**, respectively. We assume without loss of generality that $a_1 \leq a_2$ in the following.

Case (1): When $M = 2$, by the definition of $\mathbb{T}_{k,a_1,a_2,(M)}$, we know $\{j_1, j_2\} = \{j_3, j_4\}$, $\{j_5, j_6\} = \{j_7, j_8\}$, and $|\{j_t : t = 1, \ldots, 8\}| \leq 4$. Then $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(2)} = O\{\prod_{l=1}^{2} c(n, a_l) p^4 n^{a_1+a_2-3}\} = o(n^{-2})$ by the boundedness of moments in Condition 3.2.1 and the definition of $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(2)}$.

We next prove $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)} = o(n^{-2})$. Recall that we consider $|\cup_{l=1}^{4} \{\mathbf{i}^{(l)}\}| =$

a_1+a_2-2 here by the construction of $\text{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)}$. Suppose $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(2)}\}| = s$, where $s \le a_1-1$. Then symmetrically $|\{\mathbf{i}^{(3)}\}\cap\{\mathbf{i}^{(4)}\}| = s$. Further assume $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}| = s_1$, then $|\{\mathbf{i}^{(2)}\}\cap\{\mathbf{i}^{(3)}\}| = a_1-1-s-s_1$, $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(4)}\}| = a_1-1-s-s_1$ and $|\{\mathbf{i}^{(2)}\}\cap\{\mathbf{i}^{(4)}\}| = a_2-a_1+s_1$. It follows that $|(\{\mathbf{i}^{(1)}\}\cup\{\mathbf{i}^{(2)}\})\cap(\{\mathbf{i}^{(3)}\}\cup\{\mathbf{i}^{(4)}\})| = a_1+a_2-2-2s$. Note that (B.148) $=0$ if $a_1+a_2-2-2s=0$, which can only be achieved when $a_1=a_2$ and $s=a_1-1$. It remains to consider $a_1+a_2-2-2s\ge 1$, that is, $0\le s\le A_0$, where $A_0=(a_1+a_2-3)/2$. Given s and s_1, we have

$$\begin{aligned}\text{(B.148)} = {} & \Big\{\mathrm{E}\Big(\prod_{t=1,2,5,6} x_{1,j_t}\Big)\Big\}^{s_1}\Big\{\mathrm{E}\Big(\prod_{t=3,4,7,8} x_{1,j_t}\Big)\Big\}^{a_2-a_1+s_1} \\ & \times\Big\{\mathrm{E}\Big(\prod_{t=3,4,5,6} x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=1,2,7,8} x_{1,j_t}\Big)\Big\}^{a_1-1-s-s_1} \\ & \times\Big\{\mathrm{E}\Big(\prod_{t=1,2,3,4} x_{1,j_t}\Big)\mathrm{E}\Big(\prod_{t=5,6,7,8} x_{1,j_t}\Big)\Big\}^{s+1}.\end{aligned} \tag{B.150}$$

Under the considered **Case (1)**, $\{j_1,j_2\}=\{j_3,j_4\}$ and $\{j_5,j_6\}=\{j_7,j_8\}$. If $|\{j_t:t=1,\ldots,8\}|\le 3$, we know by Condition 3.2.1,

$$\Bigg|\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in J_A^c}} \text{(B.148)}\times\mathbf{1}_{|\{j_t:t=1,\ldots,8\}|\le 3}\Bigg| = O(p^3). \tag{B.151}$$

If $|\{j_t:t=1,\ldots,8\}|=4$, $\{j_1,j_2\}\cap\{j_5,j_6\}=\emptyset$. By Conditions 3.2.1, 3.2.6 and 3.2.5, we know $\mathrm{E}(\prod_{t=1}^4 x_{1,j_t})=\kappa_1\sigma_{j_1,j_1}\sigma_{j_2,j_2}=O(1)$ and similarly $\mathrm{E}(\prod_{t=5}^8 x_{1,j_t})=O(1)$. By (B.150), (B.148) $\ne 0$ only if $\mathrm{E}(\prod_{t=1,2,5,6} x_{1,j_t})\ne 0$. This induces (j_1,j_5), $(j_2,j_6)\in J_A$ or (j_1,j_6), $(j_2,j_5)\in J_A$, and then (B.148) $=O(\rho^{2(a_1+a_2-2s)})$. By the symmetricity of j indexes, we have

$$\begin{aligned}&\Bigg|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} \text{(B.148)}\times\mathbf{1}_{\{|\{j_t:t=1,\ldots,8\}|=4\}}\Bigg| \\ &\le C\sum_{(j_1,j_5),(j_2,j_6)\in J_A}\rho^{2(a_1+a_2-2-2s)} \le C|J_A|^2\rho^{2(a_1+a_2-2-2s)}.\end{aligned} \tag{B.152}$$

By (B.151) and (B.152),

$$\text{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)} = \sum_{s=0}^{A_0} O\Big\{p^3 + |J_A|^2\rho^{2(a_1+a_2-2-2s)}\Big\}n^{a_1+a_2-2}\prod_{l=1}^{2} c(n,a_l).$$

Note that $O(p^3n^{a_1+a_2-2})\prod_{l=1}^{2} c(n,a_l) = o(n^{-2})$, and

$$\begin{aligned}
&|J_A|^2\rho^{2(a_1+a_2-2-2s)}n^{a_1+a_2-2}c(n,a_1)c(n,a_2) \qquad\qquad (\text{B.153})\\
&= O(1)p^{-4}n^{-2}|J_A|^{2-\frac{2(a_1+a_2-2-2s)}{a_1+a_2}}(|J_A|\rho^{a_1}\times|J_A|\rho^{a_2})^{\frac{2(a_1+a_2-2-2s)}{a_1+a_2}}\\
&= O(1)|J_A|^{2-\frac{2(a_1+a_2-2-2s)}{a_1+a_2}}p^{\frac{2(a_1+a_2-2-2s)}{a_1+a_2}-4}n^{-(a_1+a_2-2-2s)-2}\\
&= O(1)|J_A|^{-\frac{a_1+a_2-2-2s}{a_1+a_2}}(|J_A|/p^2)^{2-\frac{a_1+a_2-2-2s}{a_1+a_2}}n^{-(a_1+a_2-2-2s)-2}\\
&= o(n^{-2}).
\end{aligned}$$

Therefore $\text{var}\{\mathbb{T}_{k,a_1,a_2,(M)}\}_{(1)} = o(n^{-2})$.

Case (2): When $M = 3$, we assume without loss of generality that $j_1 = j_3$ and $j_5 = j_7$, then

$$\{j_1,j_2,j_3,j_4\} = \{j_1,j_2,j_4\} \text{ and } \{j_5,j_6,j_7,j_8\} = \{j_5,j_6,j_8\}. \qquad (\text{B.154})$$

It follows that $\text{E}(\prod_{t=1}^{4} x_{1,j_t}) = \kappa_1\sigma_{j_1,j_1}\sigma_{j_2,j_4}$ and $\text{E}(\prod_{t=5}^{8} x_{1,j_t}) = \kappa_1\sigma_{j_5,j_5}\sigma_{j_6,j_8}$, which are 0 when (j_2,j_4) and $(j_6,j_8) \in J_A^c$; and are $O(\rho)$ when (j_2,j_4) and $(j_6,j_8) \in J_A$. This suggests that if (B.148) $\neq 0$, (j_2,j_4) and $(j_6,j_8) \in J_A$.

We first examine $\text{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(1)}$, which is the part of summation in $\text{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}$ when $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}| = a_1+a_2-2$. Recall that the two claims in (B.149) also hold here. Similarly to **Case (1)** above, we still assume $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(2)}\}| = s$, and $|\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}| = s_1$, then (B.150) holds. We next discuss several sub-cases based on the size of the set $\{j_t : t = 1,\dots,8\}$.

Case (2.1): When $|\{j_t : t = 1,\dots,8\}| = 6$, we know $\{j_1,j_2,j_3,j_4\}\cap\{j_5,j_6,j_7,j_8\} =$

$\emptyset$ by (B.154). Then by (B.150), we know if (B.148) $\neq 0$, then $(j_2, j_4), (j_6, j_8), (j_1, j_5), (j_2, j_6) \in J_A$ or $(j_2, j_4), (j_6, j_8), (j_1, j_6), (j_2, j_5) \in J_A$. Thus by the symmetricity of the j indexes, we have

$$\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{(\text{B.148})\neq 0} \times \mathbf{1}_{|\{j_t:t=1,\ldots,8\}|=6} \leq C \sum_{\substack{(j_1,j_5),(j_2,j_6),\\(j_2,j_4),(j_6,j_8)\in J_A}} 1 \leq C|J_A|^3.$$

By Conditions 3.2.6 and 3.2.5, (B.148) $= O(\rho^{\tilde{A}_1})$, where $\tilde{A}_1 = 2(a_1 + a_2 - 2 - 2s) + 2(s+1) = 2(a_1 + a_2) - 2(s+1)$. Thus

$$\left| \sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} (\text{B.148})\mathbf{1}_{|\{j_t:t=1,\ldots,8\}|=6} \right| = O(|J_A|^3 \rho^{\tilde{A}_1}).$$

Case (2.2): When $|\{j_t : t = 1, \ldots, 8\}| = 5$, recall that we assume (B.154), where $j_1 = j_3$ and $j_5 = j_7$ without loss of generality. If we further assume $j_1 = j_5$, $\{j_t : t = 1, \ldots, 8\} = \{j_1, j_2, j_4, j_6, j_8\}$. Then for (B.148) $\neq 0$, $\mathrm{E}(\prod_{t=1,2,3,4} x_{1,j_t}) \times \mathrm{E}(\prod_{t=5,6,7,8} x_{1,j_t}) \neq 0$, then $(j_2, j_4), (j_6, j_8) \in J_A$ holds. In addition, under this case, (B.148) $= O\{\rho^{(a_1+a_2-2-2s)+2(s+1)}\} = O(\rho^{a_1+a_2})$, and we have

$$\left| \sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} \mathbf{1}_{(\text{B.148})=O(\rho^{a_1+a_2}),\,|\{j_t:t=1,\ldots,8\}|=5} \right| = O(p|J_A|^2).$$

If given $j_1 = j_3$ and $j_5 = j_7$, instead, assume $j_1 \neq j_5$. We have $j_1 \neq j_2$, $j_1 \neq j_4$ and $j_1 \neq j_5$. Then for (B.148) $\neq 0$, by discussing different cases of j indexes, we know that (B.148) achieves the order between $O(\rho^{\tilde{A}_1})$ and $O(\rho^{\tilde{A}_2})$ where $\tilde{A}_1$ is defined as above and $\tilde{A}_2 = 2(s+1) + (1+2) \times (a_1 + a_2 - 2s - 2)/2 = 3(a_1 + a_2)/2 - (s+1)$. Moreover, we have

$$\left| \sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{\{(\text{B.148})=O(\rho^u),\,\tilde{A}_2\leq u\leq \tilde{A}_1,\,|\{j_t:t=1,\ldots,8\}|=5\}} \right| = O(D_{\max}|J_A|^2).$$

In summary,

$$\Bigg|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} (\text{B.148}) \times \mathbf{1}_{|\{j_t:t=1,\ldots,8\}|=5}\Bigg|$$
$$= O(D_{\max}|J_A|^2\rho^{\tilde{A}_1}) + O(D_{\max}|J_A|^2\rho^{\tilde{A}_2}) + O(p|J_A|^2\rho^{a_1+a_2}).$$

Case (2.3): When $|\{j_t : t = 1, \ldots, 8\}| = 4$, similarly as case (2.3), we can discuss $j_1 = j_5$ and $j_1 \neq j_5$ respectively. When $j_1 = j_5$, we note that (B.148) can achieve the orders between $O(\rho^{a_1+a_2})$ and $O(\rho^{\tilde{A}_3})$ with $\tilde{A}_3 = (a_1 + a_2 - 2 - 2s)/2 + 2(s + 1) = (a_1 + a_2)/2 + s + 1$. Moreover,

$$\Bigg|\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{(\text{B.148})=O(\rho^u),\tilde{A}_3\le u\le a_1+a_2,\,|\{j_t:t=1,\ldots,8\}|=4}\Bigg| = O(pD_{\max}|J_A|).$$

In addition, when $j_1 \neq j_5$, we note that (B.148) can achieve the order between $O(\rho^{a_1+a_2})$ and $O(\rho^{\tilde{A}_1})$. Under this case,

$$\Bigg|\sum_{\substack{(j_1,j_2),(j_3,j_4),\\(j_5,j_6),(j_7,j_8)\in J_A^c}} \mathbf{1}_{(\text{B.148})=O(\rho^u),\tilde{A}_4\le u\le a_1+a_2,\,|\{j_t:t=1,\ldots,8\}|=4}\Bigg| = O(|J_A|^2).$$

In summary, by $|J_A| \le pD_{\max}$,

$$\Bigg|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} (\text{B.148}) \times \mathbf{1}_{|\{j_t:t=1,\ldots,8\}|=4}\Bigg|$$
$$= O(pD_{\max}|J_A|\rho^{\tilde{A}_3}) + O(pD_{\max}|J_A|\rho^{a_1+a_2}) + O(|J_A|^2\rho^{\tilde{A}_1}).$$

Case (2.4): When $|\{j_t : t = 1, \ldots, 8\}| \le 3$, we know by Condition 3.2.1,

$$\Bigg|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} (\text{B.148}) \times \mathbf{1}_{|\{j_t:t=1,\ldots,8\}|\le 3}\Bigg| = O(p^3).$$

In summary, combining Cases (2.1)–(2.4) above, we know

$$\begin{aligned}
&\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(1)} \qquad\qquad\qquad\qquad\qquad\qquad \text{(B.155)}\\
&= \prod_{l=1}^{2} c(n,a_l) n^{a_1+a_2-2} \sum_{s=0}^{A_0} \Big\{ O(p^3) + O(|J_A|^3 \rho^{\tilde{A}_1}) \\
&\quad + O(D_{\max}|J_A|^2\rho^{\tilde{A}_1}) + O(D_{\max}|J_A|^2\rho^{\tilde{A}_2}) + O(p|J_A|^2\rho^{a_1+a_2}) \\
&\quad + O(pD_{\max}|J_A|\rho^{\tilde{A}_3}) + O(pD_{\max}|J_A|\rho^{a_1+a_2}) + O(|J_A|^2\rho^{\tilde{A}_1}) \Big\},
\end{aligned}$$

where $\tilde{A}_1 = 2(a_1+a_2)-2(s+1)$, $\tilde{A}_2 = 3(a_1+a_2)/2-(s+1)$, and $\tilde{A}_3 = (a_1+a_2)/2+s+1$.

Note that

$$\begin{aligned}
&\prod_{l=1}^{2} c(n,a_l) \times n^{a_1+a_2-2}|J_A|^3\rho^{\tilde{A}_1} \\
&= p^{-4}n^{-2}|J_A|^3\rho^{2(a_1+a_2-s-1)} \\
&= p^{-4}n^{-2}(|J_A|\rho^{a_1} \times |J_A|\rho^{a_2})^{\frac{2(a_1+a_2-s-1)}{a_1+a_2}} |J_A|^{3-\frac{4(a_1+a_2-s-1)}{a_1+a_2}} \qquad \text{(B.156)}\\
&= O(1)n^{-2}p^{\frac{4(a_1+a_2-s-1)}{a_1+a_2}-4} n^{-(a_1+a_2-s-1)}|J_A|^{-1+\frac{4(s+1)}{a_1+a_2}} \qquad \text{(B.157)}\\
&= O(1)n^{-2}p^{-\frac{4(s+1)}{a_1+a_2}} n^{-(a_1+a_2-s-1)}|J_A|^{-1+\frac{4(s+1)}{a_1+a_2}} \\
&= O(1)n^{-2}(|J_A|/p^2)^{\frac{2(s+1)}{a_1+a_2}}|J_A|^{-1+\frac{2(s+1)}{a_1+a_2}} \\
&= o(n^2),
\end{aligned}$$

where from (B.156) to (B.157), we use $|J_A|\rho^a = O(pn^{-a/2})$, and in the last equation, we use $2(s+1) \leq a_1+a_2-1$. Following similar analysis, we know that all the terms in (B.155) are $o(n^{-2})$ and $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(1)} = o(n^{-2})$.

We next examine $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(2)}$. Note that if (B.148) $\neq 0$, (j_2,j_4) and $(j_6,j_8) \in J_A$. We can discuss different cases of $\{j_1,\ldots,j_8\}$ similarly as above. Then by Conditions 3.2.5 and 3.2.6, as $\rho = O(|J_A|^{-1/a_t}p^{1/a_t}n^{-1/2})$ for $t=1,2$, we have $\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}$ (B.148) $= O(p^4)$. Given that $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| < a_1+a_2-2$

in $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(2)}$, we obtain $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(2)} = \prod_{l=1}^{2} c(n,a_l) \times O(p^4 n^{a_1+a_2-3}) = o(n^{-2})$.

In summary, we have $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\} = o(n^{-2})$.

Case (3): When $M = 4$, we consider $j_1 \neq j_2 \neq j_3 \neq j_4$ and $j_5 \neq j_6 \neq j_7 \neq j_8$ under this case. Since $\sigma_{j_1,j_2} = \sigma_{j_3,j_4} = \sigma_{j_5,j_6} = \sigma_{j_7,j_8} = 0$,

$$\mathrm{E}(x_{1,j_1}x_{1,j_2}x_{1,j_3}x_{1,j_4}) = \kappa_1(\sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3}),$$
$$\mathrm{E}(x_{1,j_5}x_{1,j_6}x_{1,j_7}x_{1,j_8}) = \kappa_1(\sigma_{j_5,j_7}\sigma_{j_6,j_8} + \sigma_{j_5,j_8}\sigma_{j_6,j_7}),$$

which are $O(\rho^2)$. Following similar analysis to **Case (2)**, we can examine the different cases when $|\{j_t : t = 1, \ldots, 8\}|$ is between 4 and 8, and obtain,

$$\begin{aligned}
&\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(4)}\}_{(1)} \qquad\qquad \text{(B.158)}\\
=\ & O(1)\prod_{l=1}^{2} c(n,a_l) \times n^{a_1+a_2-2} \sum_{s=0}^{A_0} \Big[|J_A|^2\rho^{4(s+1)}\\
&+ D_{\max}|J_A|^2\rho^{4(s+1)}\Big(\rho^{a_1-1-s} + \rho^{a_2-1-s}\Big)\\
&+ \max\{|J_A|, D_{\max}^2\} \times |J_A|^2\rho^{4(s+1)}\Big(\rho^{2(a_1-1-s)} + \rho^{2(a_2-1-s)}\Big)\\
&+ D_{\max}|J_A|^3\Big(\rho^{2(a_1+a_2)-(a_1-1-s)} + \rho^{2(a_1+a_2)-(a_2-1-s)}\Big)\\
&+ |J_A|^4\rho^{2(a_1+a_2)}\Big].
\end{aligned}$$

Note that $\prod_{l=1}^{2} c(n,a_l) n^{a_1+a_2-2}|J_A|^4\rho^{2(a_1+a_2)} = O(1)p^{-4}n^{-2}p^4 n^{-(a_1+a_2)} = o(n^{-2})$. Moreover,

$$\begin{aligned}
&\prod_{l=1}^{2} c(n,a_l) \times n^{a_1+a_2-2} D_{\max}|J_A|^2\rho^{4(s+1)}\Big(\rho^{a_1-1-s} + \rho^{a_2-1-s}\Big) \qquad \text{(B.159)}\\
=\ & p^{-4}n^{-2}D_{\max}|J_A|^2\Big(\rho^{a_1+3(s+1)} + \rho^{a_2+3(s+1)}\Big).
\end{aligned}$$

To show (B.159) $= o(n^{-2})$ by symmetricity, it suffices to show for any integer a_1,

$p^{-4}D_{\max}|J_A|^2\rho^{a_1+3(s+1)} = o(1)$.

$$\begin{aligned}
& p^{-4}D_{\max}|J_A|^2\rho^{a_1+3(s+1)} \\
= \; & p^{-4}D_{\max}(|J_A|\rho^{a_1})^{\frac{a_1+3(s+1)}{a_1}}|J_A|^{2-\frac{a_1+3(s+1)}{a_1}} && \text{(B.160)} \\
= \; & O(1)p^{-4}D_{\max}(pn^{-a_1/2})^{\frac{a_1+3(s+1)}{a_1}}|J_A|^{2-\frac{a_1+3(s+1)}{a_1}} && \text{(B.161)} \\
= \; & O(1)n^{-\frac{a_1+3(s+1)}{2}}(|J_A|/p^2)^{1-\frac{(s+1)}{a_1}} \\
& \times(D_{\max}/p)^{1-\frac{s+1}{a_1}}(D_{\max}/|J_A|)^{\frac{s+1}{a_1}}|J_A|^{-\frac{s+1}{a_1}} \\
= \; & o(1),
\end{aligned}$$

where from (B.160) to (B.161), we use $|J_A|\rho^{a_1} = O(pn^{-a_1/2})$, and in the last equation we use $|J_A| = o(p^2)$, $D_{\max} \le p$ and $D_{\max} \le |J_A|$. For other terms in (B.158), similar analysis can be applied and we have $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(4)}\}_{(1)} = o(n^{-2})$.

In addition, similarly to the analysis of $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(3)}\}_{(2)}$, by Conditions 3.2.5 and 3.2.6, we still have $\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c}$ (B.148) $= O(p^4)$. Since $|\cup_{l=1}^4 \{\mathbf{i}^{(l)}\}| < a_1 + a_2 - 2$ in $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(4)}\}_{(2)}$ by construction, we obtain $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(4)}\}_{(2)} = \prod_{l=1}^2 c(n,a_l) \times O\{p^4 n^{a_1+a_2-3}\} = o(n^{-2})$. In summary, $\mathrm{var}\{\mathbb{T}_{k,a_1,a_2,(4)}\} = o(n^{-2})$ is proved.

B.5.15 Proof of Lemma B.1.14

Similarly to Section B.5.7,

$$\sum_{k=1}^n \mathrm{E}(D_{n,k}^4) = \sum_{k=1}^n \sum_{1\le r_1,r_2,r_3,r_4\le m} \prod_{l=1}^4 t_{r_l} \times \mathrm{E}\Big(\prod_{l=1}^4 A_{n,k,a_{r_l}}\Big),$$

where we use the redefined notation in Section B.1.4. To prove Lemma B.1.6, it suffices to show that for given $1 \le k \le n$ and $1 \le r_1, r_2, r_3, r_4 \le m$, we have $\mathrm{E}(\prod_{l=1}^4 A_{n,k,a_{r_l}}) = o(n^{-1})$. Moreover by the Cauchy-Schwarz inequality, it suffices to show $\mathrm{E}(A_{n,k,a}^4) = o(n^{-1})$ for $a \in \{a_1, \ldots, a_m\}$. Following (B.92), we have $A_{n,k,a} = 0$

when $k < a$; and when $k \geq a$,

$$\mathrm{E}(A_{n,k,a}^4) = c^2(n,a) \sum_{\substack{\mathbf{i}^{(l)} \in \mathcal{P}(k-1,a-1),\, l=1,2,3,4; \\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in J_A^c}} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8),$$

where $\mathbf{i}^{(l)} = (i_1^{(l)}, \ldots, i_a^{(l)})$ represents tuples $1 \leq i_1^{(l)} \neq \ldots \neq i_a^{(l)} \leq n$, and

$$Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = \mathrm{E}\Big(\prod_{r=1}^{8} x_{k,j_r}\Big)\, \mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a-1} x_{i_t^{(l)}, j_{2l-1}} x_{i_t^{(l)}, j_{2l}}\Big).$$

As $c(n,a) = \Theta(p^{-1}n^{-a/2})$, to prove $\mathrm{E}(A_{n,k,a}^4) = o(n^{-1})$, it suffices to show

$$\sum_{\substack{\mathbf{i}^{(l)} \in \mathcal{P}(k-1,a-1),\, l=1,2,3,4; \\ (j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in J_A^c}} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = o(p^4 n^{2a-1}).$$

Since $\sigma_{j_1,j_2} = 0$ if $(j_1,j_2) \in J_A^c$, then similarly to Section B.5.7, we know that $Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) \neq 0$ only when $|\bigcup_{l=1}^{4}\{\mathbf{i}^{(l)}\}| \leq 2(a-1)$, and similarly to (B.96),

$$\sum_{\mathbf{i}^{(l)} \in \mathcal{P}(k-1,a-1),\, l=1,\ldots,4} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = O(n^{2a-2}).$$

It then remains to show

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8) \in J_A^c} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) = O(p^4). \tag{B.162}$$

We next prove by discussing $|\{j_t : t = 1, \ldots, 8\}|$ and the corresponding value of $Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8)$. By Condition 3.2.6, $Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8)$ can be written as certain linear combination of $\prod_{t=1}^{4a}(\sigma_{j_{g_{2t-1}}, j_{g_{2t}}})$, where $g_{2t-1} \neq g_{2t}$ and $(g_1, \ldots, g_{8a})$ contain a number of $1, \ldots, 8$ respectively. If $|\{j_t : t = 1, \ldots, 8\}| \leq 4$, by Condition

3.2.1,

$$\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\times \mathbf{1}_{\{|\{j_t:t=1,\ldots,8\}|\le 4\}} = O(p^4).$$

If $|\{j_t : t = 1,\ldots,8\}| = 5$, note that for $j_1 \neq j_2$, $\sigma_{j_1,j_2} \neq 0$ only when $(j_1, j_2) \in J_A$, then

$$\begin{aligned}&\Big|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\times \mathbf{1}_{\{|\{j_t:t=1,\ldots,8\}|=5\}}\Big|\\ \le\ & C\sum_{\substack{1\le j_1,j_2,j_5\le p,\\ (j_6,j_8)\in J_A}} \sigma^a_{j_1,j_1}\sigma^a_{j_2,j_2}\sigma^a_{j_5,j_5}\sigma^a_{j_6,j_8} = O(p^3|J_A|\rho^a) = o(p^4),\end{aligned}$$

where in the last equation, we use $|J_A|\rho^a = O(pn^{-a/2})$. In addition, similarly, if $|\{j_t : t = 1,\ldots,8\}| = 6$,

$$\begin{aligned}&\Big|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\times \mathbf{1}_{\{|\{j_t:t=1,\ldots,8\}|=6\}}\Big|\\ \le\ & C\sum_{\substack{1\le j_1,j_2\le p,\\ (j_5,j_7),(j_6,j_8)\in J_A}} \sigma^a_{j_1,j_1}\sigma^a_{j_2,j_2}\sigma^a_{j_5,j_7}\sigma^a_{j_6,j_8} = O(p^2|J_A|^2\rho^{2a}) = o(p^4).\end{aligned}$$

If $|\{j_t : t = 1,\ldots,8\}| = 7$,

$$\begin{aligned}&\Big|\sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)},\mathbf{j}_8)\times \mathbf{1}_{\{|\{j_t:t=1,\ldots,8\}|=7\}}\Big|\\ \le\ & C\sum_{\substack{1\le j_1\le p,\\ (j_2,j_4),(j_5,j_7),(j_6,j_8)\in J_A}} \sigma^a_{j_1,j_1}\sigma^a_{j_2,j_4}\sigma^a_{j_5,j_7}\sigma^a_{j_6,j_8} = O(p|J_A|^3\rho^{3a}) = o(p^4).\end{aligned}$$

If $|\{j_t : t = 1, \ldots, 8\}| = 8$,

$$
\begin{aligned}
&\Big| \sum_{(j_1,j_2),(j_3,j_4),(j_5,j_6),(j_7,j_8)\in J_A^c} Q^*(\mathbf{i}^{(1)}, \mathbf{i}^{(2)}, \mathbf{i}^{(3)}, \mathbf{i}^{(4)}, \mathbf{j}_8) \times \mathbf{1}_{\{|\{j_t:t=1,\ldots,8\}|=8\}} \Big| \\
&\leq C \sum_{(j_1,j_3),(j_2,j_4),(j_5,j_7),(j_6,j_8)\in J_A} \sigma^a_{j_1,j_3}\sigma^a_{j_2,j_4}\sigma^a_{j_5,j_7}\sigma^a_{j_6,j_8} = O(|J_A|^4\rho^{4a}) = o(p^4).
\end{aligned}
$$

In summary, (B.162) is obtained and Lemma B.1.14 is proved.

B.5.16 Proof of Lemma B.2.1

We prove $\mathrm{var}(\sum_{k=1}^n \pi^2_{n,k}) \to 0$ and $\sum_{k=1}^n \mathrm{E}(D^4_{n,k}) \to 0$ in the following Sections B.5.16.1 and B.5.16.2, respectively.

B.5.16.1 Proof of $\mathrm{var}(\sum_{k=1}^n \pi^2_{n,k}) \to 0$.

Similarly to Section B.5.6, $D_{n,k} = \sum_{r=1}^m t_r A_{n,k,a_r}$, and then $\pi^2_{n,k} = \sum_{1\leq r_1,r_2\leq m} t_{r_1}t_{r_2}\times \mathrm{E}_{k-1}(A_{n,k,a_{r_1}}A_{n,k,a_{r_2}})$. Note that by the Cauchy-Schwarz inequality, for some constant C, $\mathrm{var}(\sum_{k=1}^n \pi^2_{n,k}) \leq Cn^2 \max_{1\leq k\leq n;\, 1\leq r_1,r_2\leq m} \mathrm{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}})$, where $c(n,a) = [a\times \{\sigma(a)P^n_a\}^{-1}]^2$ and for two integers a_1 and a_2 we still define $\mathbb{T}_{k,a_1,a_2} = \mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})$. In particular, when $k < \max\{a_1,a_2\}$, $\mathbb{T}_{k,a_1,a_2} = 0$; when $k \geq \max\{a_1,a_2\}$,

$$
\begin{aligned}
\mathbb{T}_{k,a_1,a_2} &= \mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2}) \\
&= \sum_{\substack{1\leq j_1,j_2\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2}} \{c(n,a_1)c(n,a_2)\}^{1/2}\sigma_{j_1,j_2}\prod_{l=1}^{2}\prod_{t=1}^{a_l-1} x_{i^{(l)}_t,j_l}.
\end{aligned}
$$

To prove Lemma $\mathrm{var}(\sum_{k=1}^n \pi^2_{n,k}) \to 0$, it suffices to prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$, where $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = \mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2$. We consider without loss of generality that $k \geq \max\{a_1,a_2\}$.

When $\{\mathbf{i}^{(1)}\} \neq \{\mathbf{i}^{(2)}\}$, $\mathrm{E}(\prod_{l=1}^2\prod_{t=1}^{a_l} x_{i^{(l)}_t,j_t}) = 0$; and when $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$, it induces $a_1 = a_2$ and $\mathrm{E}(\prod_{l=1}^2\prod_{t=1}^{a_l} x_{i^{(l)}_t,j_t}) = \sigma^a_{j_1,j_2}$ where we write $a_1 = a_2 = a$. It follows that when $a_1 \neq a_2$, $\mathrm{E}(\mathbb{T}_{k,a_1,a_2}) = 0$; when $a_1 = a_2 = a$,

$$\mathrm{E}(\mathbb{T}_{k,a_1,a_2}) = \sum_{\substack{1\le j_1,j_2\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2}} \mathbf{1}_{\{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}\}} \times \{c(n,a_1)c(n,a_2)\}^{1/2}\sigma_{j_1,j_2}^{a}.$$

Then

$$\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 = \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}\\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}}\right\}} \prod_{l=1}^{2} c(n,a_l) \times (\sigma_{j_1,j_2}\sigma_{j_3,j_4})^{a}.$$

In addition, we obtain

$$\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) = \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2,3,4}} \Big\{\prod_{l=1}^{2} c(n,a_l)\sigma_{j_{2l-1},j_{2l}}\Big\}\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_l}\Big),$$

where for simplicity of representation, we set $a_3 = a_1$ and $a_4 = a_2$. Define

$$\begin{aligned} G_{k,a_1,a_2,1} = &\sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\},\\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\},\\ \{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}=\emptyset}\right\}} \\ &\times \Big\{\prod_{l=1}^{2} c(n,a_l)\sigma_{j_{2l-1},j_{2l}}\Big\}\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_l}\Big). \end{aligned}$$

Since $|\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2| \le |\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - G_{k,a_1,a_2,1}| + |G_{k,a_1,a_2,1} - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2|$, we next prove $\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - G_{k,a_1,a_2,1} = o(n^{-2})$ and $G_{k,a_1,a_2,1} - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 = o(n^{-2})$ respectively.

Step I: $\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - G_{k,a_1,a_2,1} = o(n^{-2})$ When $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$, $\{\mathbf{i}^{(3)}\} = \{\mathbf{i}^{(4)}\}$, and $\{\mathbf{i}^{(1)}\} \cap \{\mathbf{i}^{(3)}\} = \emptyset$, it implies that $a_1 = a_2 = a$, $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}| \le a_1 + a_2 - 3$, and

$$\Big(\prod_{l=1}^{2}\sigma_{j_{2l-1},j_{2l}}\Big) \times \mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_l}\Big) = (\sigma_{j_1,j_2}\sigma_{j_3,j_4})^{a}.$$

It follows that if $a_1 \neq a_2$, $\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{k,a_1,a_2,1} = 0$; if $a_1 = a_2 = a$,

$$\begin{aligned}
&\Big|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{k,a_1,a_2,1}\Big| \\
&= c(n,a_1)c(n,a_2)O(n^{a_1+a_2-3})\Big|\sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a\Big| = o(n^{-2})
\end{aligned}$$

where we use $c(n,a) = \Theta(p^{-1}n^{-a})$ and by Condition 3.3.1,

$$\sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4})^a = O(p^2). \tag{B.163}$$

Step II: $G_{k,a_1,a_2,1} - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 = o(n^{-2})$ We write $\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{k,a_1,a_2,1} = G_{k,a_1,a_2,2} + G_{k,a_1,a_2,3}$, where

$$\begin{aligned}
G_{k,a_1,a_2,2} = &\sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\},\\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\},\\ \{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}\neq\emptyset}\right\}} \\
&\times\Big\{\prod_{l=1}^{2}c(n,a_l)\sigma_{j_{2l-1},j_{2l}}\Big\}\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1}x_{i_t^{(l)},j_l}\Big),
\end{aligned}$$

and

$$\begin{aligned}
G_{k,a_1,a_2,3} = &\sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1):\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}\neq\{\mathbf{i}^{(2)}\}\text{ or}\\ \{\mathbf{i}^{(3)}\}\neq\{\mathbf{i}^{(4)}\}}\right\}} \\
&\times\Big\{\prod_{l=1}^{2}c(n,a_l)\sigma_{j_{2l-1},j_{2l}}\Big\}\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1}x_{i_t^{(l)},j_l}\Big).
\end{aligned}$$

For $G_{k,a_1,a_2,2}$, it is a summation over the indexes satisfying $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}, \{\mathbf{i}^{(3)}\} = \{\mathbf{i}^{(4)}\}$ and $\{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\} \neq \emptyset$. Thus $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| \le a_1+a_2-3$, and by $c(n,a) = \Theta(p^{-1}n^{-a})$ and (B.163),

$$|G_{k,a_1,a_2,2}| \le Cp^{-2}n^{-(a_1+a_2)}n^{a_1+a_2-3}\sum_{1\le j_1,j_2,j_3,j_4\le p}\sigma_{j_1,j_2}\sigma_{j_3,j_4} = o(n^{-2}).$$

For $G_{k,a_1,a_2,3}$, it is a summation over the indexes satisfying $\{\mathbf{i}^{(1)}\} \neq \{\mathbf{i}^{(2)}\}$ or $\{\mathbf{i}^{(3)}\} \neq \{\mathbf{i}^{(4)}\}$. We assume without loss of generality that $\{\mathbf{i}^{(1)}\} \neq \{\mathbf{i}^{(2)}\}$ and there exists an index $m \in \{\mathbf{i}^{(1)}\}$ but $m \notin \{\mathbf{i}^{(2)}\}$. Similarly to Section B.5.6, we know

$$\Big(\prod_{l=1}^{2} \sigma_{j_{2l-1},j_{2l}}\Big) \times \mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t=1}^{a_l-1} x_{i_t^{(l)},j_l}\Big) \tag{B.164}$$

is nonzero only when $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}| \leq a_1+a_2-2$, that is, each index appears at least twice among the four sets $\{\mathbf{i}^{(l)}\}, l=1,2,3,4$. Therefore, we know if (B.164) $\neq 0$, $m \in \{\mathbf{i}^{(3)}\} \cup \{\mathbf{i}^{(4)}\}$. If $m \in \{\mathbf{i}^{(3)}\}$ but $m \notin \{\mathbf{i}^{(4)}\}$, (B.164) $= \sigma_{j_1,j_2}\sigma_{j_3,j_4}\sigma_{j_1,j_3}\mathrm{E}(\text{other terms})$. Under this case, we define $\tilde{K}_0 = -(2+\epsilon)(4+\gamma)\log p/(\epsilon \log\delta)$, where γ and ϵ are some positive constants and δ is from Condition 3.3.1. Then we have

$$\begin{aligned}\sum_{1\leq j_1,j_2,j_3,j_4\leq p} (\text{B.164}) &\leq C \sum_{1\leq j_1,j_2,j_3,j_4\leq p} \sigma_{j_1,j_2}\sigma_{j_3,j_4}\sigma_{j_1,j_3} \\ &\leq C \sum_{\substack{|j_1-j_2|\leq \tilde{K}_0,\\ |j_3-j_4|\leq \tilde{K}_0,\\ |j_1-j_3|\leq \tilde{K}_0}} 1 + C\sum_{|j_1-j_2|\geq \tilde{K}_0} \delta^{|j_1-j_2|\epsilon/(2+\epsilon)} \\ &= O(p\tilde{K}_0^2) + O(p^4 p^{-(4+\gamma)}),\end{aligned} \tag{B.165}$$

where in the second inequality, we use the symmetricity of j indexes and also use Lemma B.5.1 similarly as in Section B.2.1. If $m \in \{\mathbf{i}^{(4)}\}$ but $m \notin \{\mathbf{i}^{(s)}\}$, (B.165) also holds similarly. If $m \in \{\mathbf{i}^{(3)}\}$ and $m \in \{\mathbf{i}^{(4)}\}$, (B.164) $= \sigma_{j_1,j_2}\sigma_{j_3,j_4}\mathrm{E}(x_{m,j_1}x_{m,j_3}x_{m,j_4}) \times \mathrm{E}(\text{other terms})$. Similarly to (B.165), as $\mathrm{E}(\mathbf{x}) = \mathbf{0}$, if $|j_1-j_3| > \tilde{K}_0$ and $|j_1-j_4| > \tilde{K}_0$, (B.164) $\leq C\delta^{|j_1-j_2|\epsilon/(2+\epsilon)}$. Thus under this case, we also have $\sum_{1\leq j_1,j_2,j_3,j_4\leq p}$ (B.164) $= O(p\tilde{K}_0^2) + O(p^{-\gamma})$. Recall that (B.164) $\neq 0$ only when $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}| \leq a_1+a_2-2$. By

$c(n,a)=\Theta(p^{-1}n^{-a})$ and $\tilde{K}_0=O(\log p)$,

$$\begin{aligned}|G_{k,a_1,a_2,3}| &\le Cp^{-2}n^{-(a_1+a_2)}n^{a_1+a_2-2}\sum_{1\le j_1,j_2,j_3,j_4\le p}|(\mathrm{B}.164)| \\ &= n^{-2}p^{-2}\Big\{O(p\tilde{K}_0^2)+O(p^{-\gamma})\Big\}=o(n^{-2}).\end{aligned}$$

In summary,

$$\mathrm{var}(\mathbb{T}_{k,a_1,a_2})\le|\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2})-G_{k,a_1,a_2,1}|+|G_{k,a_1,a_2,2}|+|G_{k,a_1,a_2,3}|=o(n^{-2}),$$

and then $\mathrm{var}(\sum_{k=1}^n\pi^2_{n,k})\to 0$ is proved.

B.5.16.2 Proof of $\sum_{k=1}^n \mathrm{E}(D^4_{n,k})\to 0$.

Similarly to Section B.5.7,

$$\sum_{k=1}^n \mathrm{E}(D^4_{n,k})=\sum_{k=1}^n\sum_{1\le r_1,r_2,r_3,r_4\le m}\prod_{l=1}^4 t_{r_l}\times\mathrm{E}\Big(\prod_{l=1}^4 A_{n,k,a_{r_l}}\Big).$$

To prove $\sum_{k=1}^n \mathrm{E}(D^4_{n,k})\to 0$, it suffices to show that for given $1\le k\le n$ and finite integers (a_1,a_2,a_3,a_4), we have $\mathrm{E}(\prod_{l=1}^4 A_{n,k,a_l})=o(n^{-1})$.

In particular,

$$\mathrm{E}\Big(\prod_{l=1}^4 A_{n,k,a_l}\Big)=\Big\{\prod_{l=1}^4 c(n,a_l)\Big\}^{1/2}\sum_{\substack{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\,l=1,\ldots,4;\\ 1\le j_1,j_2,j_3,j_4\le p}}\mathrm{E}\Big(\prod_{l=1}^4 x_{k,j_l}\Big)\mathrm{E}\Big(\prod_{l=1}^4\prod_{t=1}^{a_l-1}x_{i_t,j_l}\Big).$$

Similarly to Section B.5.7, we have $\mathrm{E}(\prod_{l=1}^4\prod_{t=1}^{a_l-1}x_{i_t,j_l})\neq 0$ only when $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|\le\sum_{l=1}^4(a_l-1)/2$. We will prove that

$$\sum_{1\le j_1,j_2,j_3,j_4\le p}\mathrm{E}\Big(\prod_{l=1}^4 x_{k,j_l}\Big)=O(p^2). \tag{B.166}$$

Then as $c(n,a) = \Theta(p^{-1}n^{-a})$,

$$\mathrm{E}\Big(\prod_{l=1}^{4} A_{n,k,a_l}\Big) = O(1)p^{-2}n^{-\sum_{l=1}^{4} a_l/2}n^{\sum_{l=1}^{4}(a_l-1)/2}p^2 = o(n^{-1}).$$

To finish the proof, it remains to show (B.166). When $|\{j_1, j_2, j_3, j_4\}| \leq 2$,

$$\sum_{1\leq j_1,j_2,j_3,j_4\leq p} \mathrm{E}\Big(\prod_{l=1}^{4} x_{k,j_l}\Big)\mathbf{1}_{\{|\{j_1,j_2,j_3,j_4\}|\leq 2\}} = O(p^2).$$

When $|\{j_1, j_2, j_3, j_4\}| \geq 3$, we assume without loss of generality that $j_1 \leq j_2 \leq j_3 \leq j_4$. For $\tilde{K}_0$ defined in Section B.5.16.1, if $|j_1 - j_2| > \tilde{K}_0$ or $|j_3 - j_4| > \tilde{K}_0$, $|\mathrm{E}(\prod_{l=1}^{4} x_{k,j_l})| \leq C\delta^{|j_1-j_2|\epsilon/(2+\epsilon)} = O(p^{-(4+\gamma)})$. If $|j_1 - j_2| \leq \tilde{K}_0$ and $|j_3 - j_4| \leq \tilde{K}_0$, but $|j_2 - j_3| > K_0$, by Lemma B.5.1,

$$\Big|\mathrm{E}\Big(\prod_{l=1}^{4} x_{k,j_l}\Big)\Big| \leq \sigma_{j_1,j_2}\sigma_{j_3,j_4} + C\delta^{|j_1-j_2|\epsilon/(2+\epsilon)} = \sigma_{j_1,j_2}\sigma_{j_3,j_4} + O(p^{-(4+\gamma)}).$$

Therefore

$$\begin{aligned}
&\sum_{1\leq j_1,j_2,j_3,j_4\leq p} \mathrm{E}\Big(\prod_{l=1}^{4} x_{k,j_l}\Big)\mathbf{1}_{\{|\{j_1,j_2,j_3,j_4\}|\geq 3\}} \\
&= O(p\tilde{K}_0^3) + O(p^4 p^{-(4+\gamma)}) + \sum_{1\leq j_1,j_2,j_3,j_4\leq p} \sigma_{j_1,j_2}\sigma_{j_3,j_4} = O(p^2),
\end{aligned}$$

where in the last equation, we use Condition 3.3.1 (2). In summary, (B.166) is proved and the proof is finished.

B.5.17 Proof of Lemma B.2.2

Under $H_0 : \boldsymbol{\mu} = \boldsymbol{\nu}$, we assume $\boldsymbol{\mu} = \boldsymbol{\nu} = \mathbf{0}$ without loss of generality by Proposition B.2.1. To derive $\mathrm{var}\{\mathcal{U}(a)\}$, we write $\mathcal{U}(a) = \sum_{j=1}^{p} \mathcal{U}^{(j)}(a)$, where we define $G(a, c) =$

$(-1)^{a-c}\binom{a}{c}(P_c^{n_x})^{-1}(P_{a-c}^{n_y})^{-1}$, and

$$\mathcal{U}^{(j)}(a) = \sum_{c=0}^{a} G(a,c) \sum_{\substack{\mathbf{k}\in\mathcal{P}(n_x,c),\\ \mathbf{s}\in\mathcal{P}(n_y,a-c)}} \prod_{t=1}^{c} x_{k_t,j} \prod_{m=1}^{a-c} y_{s_m,j}. \tag{B.167}$$

Since $\mathrm{E}\{\mathcal{U}(a)\} = 0$ under H_0,

$$\mathrm{var}\{\mathcal{U}(a)\} = \mathrm{E}\{\mathcal{U}^2(a)\} = \sum_{1\le j_1,j_2\le p} \mathrm{E}\{\mathcal{U}^{(j_1)}(a) \times \mathcal{U}^{(j_2)}(a)\}. \tag{B.168}$$

Note that for given $1 \le j_1, j_2 \le p$,

$$\mathrm{E}\{\mathcal{U}^{(j_1)}(a)\mathcal{U}^{(j_2)}(a)\} = \sum_{\substack{0\le c\le a,\\ \mathbf{k}\in\mathcal{P}(n_x,c),\\ \mathbf{s}\in\mathcal{P}(n_y,a-c)}} \sum_{\substack{0\le \tilde{c}\le a,\\ \tilde{\mathbf{k}}\in\mathcal{P}(n_x,\tilde{c}),\\ \tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-\tilde{c})}} G(a,c)G(a,\tilde{c})Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}).$$

where we define

$$Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}) = \mathrm{E}\Big(\prod_{t=1}^{c} x_{k_t,j_1} \prod_{\tilde{t}=1}^{\tilde{c}} x_{\tilde{k}_{\tilde{t}},j_2}\Big)\mathrm{E}\Big(\prod_{m=1}^{a-c} y_{s_m,j_1} \prod_{\tilde{m}=1}^{a-\tilde{c}} y_{\tilde{s}_{\tilde{m}},j_2}\Big).$$

Since we assume the $n = n_x + n_y$ copies are independent from each other and $\boldsymbol{\mu} = \boldsymbol{\nu} = \mathbf{0}$, then $Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}}) = 0$ if $\{\mathbf{k}\} \neq \{\tilde{\mathbf{k}}\}$ or $\{\mathbf{s}\} \neq \{\tilde{\mathbf{s}}\}$. If $\{\mathbf{k}\} = \{\tilde{\mathbf{k}}\}$ and $\{\mathbf{s}\} = \{\tilde{\mathbf{s}}\}$, it induces $c = \tilde{c}$ and $Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}) = \sigma_{x,j_1,j_2}^{c}\sigma_{y,j_1,j_2}^{a-c}$. It follows that

$$\begin{aligned}\mathrm{E}\{\mathcal{U}^{(j_1)}(a)\mathcal{U}^{(j_2)}(a)\} &= \sum_{c=0}^{a} G^2(c)P_c^{n_x}P_{a-c}^{n_y}c!(a-c)!\sigma_{x,j_1,j_2}^{c}\sigma_{y,j_1,j_2}^{a-c} \\ &= a!\sum_{c=0}^{a}\binom{a}{c}(P_c^{n_x})^{-1}(P_{a-c}^{n_y})^{-1}\sigma_{x,j_1,j_2}^{c}\sigma_{y,j_1,j_2}^{a-c} \\ &\simeq a!\Big(\frac{\sigma_{x,j_1,j_2}}{n_x} + \frac{\sigma_{y,j_1,j_2}}{n_y}\Big)^{a}.\end{aligned} \tag{B.169}$$

Combining (B.168) and (B.169), we obtain $\mathrm{var}\{\mathcal{U}(a)\}$. By Condition 3.3.3, we have $\mathrm{var}\{\mathcal{U}(a)\} = \Theta(pn^{-a})$.

B.5.18 Proof of Lemma B.2.3

Since under H_0, $\mathrm{E}\{\mathcal{U}(a)\} = \mathrm{E}\{\mathcal{U}(b)\} = 0$, we have $\mathrm{cov}\{\mathcal{U}(a),\mathcal{U}(b)\} = \mathrm{E}\{\mathcal{U}(a) \times \mathcal{U}(b)\}$. Following (B.167),

$$\mathrm{E}\{\mathcal{U}(a) \times \mathcal{U}(b)\} = \sum_{1\le j_1,j_2\le p} \mathrm{E}\{\mathcal{U}^{(j_1)}(a) \times \mathcal{U}^{(j_2)}(b)\}, \tag{B.170}$$

where

$$\begin{aligned}\mathrm{E}\{\mathcal{U}^{(j_1)}(a) \times \mathcal{U}^{(j_2)}(b)\} = & \sum_{\substack{0\le c\le a,\\ \mathbf{k}\in\mathcal{P}(n_x,c),\\ \mathbf{s}\in\mathcal{P}(n_y,a-c)}} \sum_{\substack{0\le \tilde{c}\le b,\\ \tilde{\mathbf{k}}\in\mathcal{P}(n_x,c),\\ \tilde{\mathbf{s}}\in\mathcal{P}(n_y,b-c)}} G(a,c)G(b,\tilde{c}) \\ & \times \mathrm{E}\Big(\prod_{t=1}^{c} x_{k_t,j_1} \prod_{\tilde{t}=1}^{\tilde{c}} x_{\tilde{k}_{\tilde{t}},j_2}\Big)\mathrm{E}\Big(\prod_{m=1}^{a-c} y_{s_m,j_1} \prod_{\tilde{m}=1}^{b-\tilde{c}} y_{\tilde{s}_{\tilde{m}},j_2}\Big).\end{aligned}$$

As $a \neq b$, $\{\mathbf{k}\} \neq \{\tilde{\mathbf{k}}\}$ and $\{\mathbf{s}\} \neq \{\tilde{\mathbf{s}}\}$ always hold. Then as $\boldsymbol{\mu} = \boldsymbol{\nu} = \mathbf{0}$, $\mathrm{E}(\prod_{t=1}^{c} x_{k_t,j_1} \times \prod_{\tilde{t}=1}^{\tilde{c}} x_{\tilde{k}_{\tilde{t}},j_2}) = 0$ and $\mathrm{E}(\prod_{m=1}^{a-c} y_{s_m,j_1} \prod_{\tilde{m}=1}^{b-\tilde{c}} y_{\tilde{s}_{\tilde{m}},j_2}) = 0$, similarly to Section B.5.3. It follows that (B.170) = 0 and the lemma is proved.

B.5.19 Proof of Lemma B.2.4

By the Cramér-Wold Theorem, to prove the asymptotic joint normality of the U-statistics, it suffices to prove that any of their fixed converges to normal. For illustration, we first prove the asymptotic normality for each $\mathcal{U}(a)$ of finite a. The similar arguments can be applied to the linear combination of finite U-statistics and then the joint normality is obtained.

Recall $\mathcal{U}(a) = \sum_{j=1}^{p} \mathcal{U}^{(j)}(a)$ from (B.167). To derive the limiting distribution of $\mathcal{U}(a)$, we use Bernstein's block method in (Ibragimov and Linnik, 1971, page 338) (also see Chen et al., 2019a; Xu et al., 2016). Specifically, we partition the sequence, $\sigma^{-1}(a) \times \mathcal{U}^{(j)}(a)$, $j = 1,\dots,p$, into r blocks, where each block contains b variables such that $rb \le p < (r+1)b$. For each $1 \le k \le r$, we partition the kth block into two sub-blocks with a larger one $A_{k,1}$ and a smaller one $A_{k,2}$. Suppose each $A_{k,1}$ has b_1 variables and each $A_{k,2}$ has $b_2 = b - b_1$ variables. We require $r \to \infty$, $b_1 \to \infty$, $b_2 \to \infty$, $rb_1/p \to 1$ and $rb_2/p \to 0$ as $p \to \infty$. We write

$$A_{k,1}(a) = \sum_{i=1}^{b_1} \mathcal{U}^{(k-1)b+i}(a), \quad A_{k,2}(a) = \sum_{i=1}^{b_2} \mathcal{U}^{(k-1)b+b_1+i}(a),$$

and further define $\mathcal{U}_1 = \sigma^{-1}(a) \sum_{k=1}^{r} A_{k,1}(a)$, $\mathcal{U}_2 = \sigma^{-1}(a) \sum_{k=1}^{r} A_{k,2}(a)$, and $\mathcal{U}_3 = \sigma^{-1}(a) \sum_{j=rb+1}^{p} \mathcal{U}^{(j)}(a)$. Thus we have the decomposition: $\sigma^{-1}(a) \times \mathcal{U}(a) = \mathcal{U}_1 + \mathcal{U}_2 + \mathcal{U}_3$.

The Bernstein's block method makes $A_{k,1}$ "almost" independent, thus the study of $\mathcal{U}_1$ may be related to the cases of sums of independent random variables. In addition, since b_2 is small compared with b_1, we will show that the sums $\mathcal{U}_2$ and $\mathcal{U}_3$ will be small compared with the total sum of variables in the sequence, i.e., $\sigma^{-1}(a) \times \mathcal{U}(a)$. In particular, we first show $\sigma^{-1}(a) \times \mathcal{U}(a) = \mathcal{U}_1 + o_p(1)$, where $o_p(1)$ represents that the remaining term converges to 0 in probability. Since $\mathrm{E}(\mathcal{U}_2) = \mathrm{E}(\mathcal{U}_3) = 0$, it suffices to prove that $\mathrm{var}(\mathcal{U}_2) = \mathrm{var}(\mathcal{U}_3) = o(1)$.

For $\mathcal{U}_2$, note that $\mathcal{U}_2 = \sigma^{-1}(a) \sum_{k=1}^{r} A_{k,2}(a)$. Then

$$\mathrm{var}(\mathcal{U}_2) \leq \sigma^{-2}(a) \sum_{\substack{1 \leq k_1, k_2 \leq r; \\ 1 \leq i_1, i_2 \leq b_2}} \left| \mathrm{cov}\Big\{ \mathcal{U}^{((k_1-1)b+b_1+i_1)}(a),\ \mathcal{U}^{((k_2-1)b+b_1+i_2)}(a) \Big\} \right|. \quad \text{(B.171)}$$

Recall $\alpha_x(s)$ and $\alpha_y(s)$ in Condition 3.3.3. Define $\alpha(s) = \alpha_x(s) + \alpha_y(s)$, then $\alpha(s) \leq C\delta^s$, where $\delta = \max\{\delta_x, \delta_y\} \in (0,1)$. By the α-mixing inequality in Lemma B.5.1,

$$\left| \mathrm{cov}\left\{ n^{a/2}\mathcal{U}^{(i)}(a), n^{a/2}\mathcal{U}^{(j)}(a) \right\} \right| \leq 8\{\alpha(|i-j|)\}^{\frac{\epsilon}{2+\epsilon}} \max_{1 \leq j \leq p} \left[\mathrm{E} \left| n^{a/2}\mathcal{U}^{(j)}(a) \right|^{2+\epsilon} \right]^{\frac{2}{2+\epsilon}}.$$

Take $\epsilon = 2$, and by Lemma B.5.8 on Page 395, we have $\max_{1 \leq j \leq p} \mathrm{E}\{n^{a/2}\mathcal{U}^{(j)}(a)\}^{2+\epsilon} <$

∞. It follows that

$$\begin{aligned}
& \left|\operatorname{cov}\left\{\mathcal{U}^{((k_1-1)b+b_1+i_1)}(a), \mathcal{U}^{((k_2-1)b+b_1+i_2)}(a)\right\}\right| \qquad \text{(B.172)}\\
= \ & n^{-a}\left|\operatorname{cov}\left\{n^{a/2}\mathcal{U}^{((k_1-1)b+b_1+i_1)}(a), n^{a/2}\mathcal{U}^{((k_2-1)b+b_1+i_2)}(a)\right\}\right| \\
\leq \ & Cn^{-a}\alpha\left\{|((k_1-1)b+b_1+i_1)-((k_2-1)b+b_1+i_2)|\right\}^{\frac{2}{4}} \\
\leq \ & Cn^{-a}\delta^{|k_1b+i_1-k_2b-i_2|/2}.
\end{aligned}$$

By (B.171), (B.172) and $\sigma^2(a) = \Theta(pn^{-a})$ from Lemma B.2.2,

$$\begin{aligned}
\operatorname{var}(\mathcal{U}_2) \ & \leq \ \sigma^{-2}(a) \sum_{\substack{1\leq k_1,k_2\leq r;\\ 1\leq i_1,i_2\leq b_2}} \left|\operatorname{cov}\left\{\mathcal{U}^{((k_1-1)b+b_1+i_1)}(a), \mathcal{U}^{((k_2-1)b+b_1+i_2)}(a)\right\}\right| \\
& \leq \ \sigma^{-2}(a) \sum_{\substack{1\leq k_1,k_2\leq r;\\ 1\leq i_1,i_2\leq b_2}} n^{-a}C\delta^{|k_1b+i_1-k_2b-i_2|/2} \\
& = \ O(1)p^{-1}n^a rb_2 n^{-a} = O(1)rb_2p^{-1},
\end{aligned}$$

which converges to 0 by our construction, i.e., $rb_2/p \to 0$. This shows that $\operatorname{var}(\mathcal{U}_2) = o(1)$. Next we exmaine $\mathcal{U}_3 = \sigma^{-1}(a)\sum_{j=rb+1}^{p}\mathcal{U}^{(j)}(a)$. Similarly, by Lemmas B.5.1 and B.5.8, and $\epsilon = 2$,

$$\begin{aligned}
\operatorname{var}(\mathcal{U}_3) \ & = \ \sigma^{-2}(a)n^{-a}\sum_{i=rb+1}^{p}\sum_{j=rb+1}^{p}\operatorname{cov}\left\{n^{a/2}\mathcal{U}^{(i)}(a), n^{a/2}\mathcal{U}^{(j)}(a)\right\} \\
& \leq \ O(1)p^{-1}n^a n^{-a}\sum_{i=rb+1}^{p}\sum_{j=rb+1}^{p} C\alpha\left(|i-j|\right)^{\frac{\epsilon}{2+\epsilon}} \\
& \leq \ O(1)p^{-1}(p-rb-1) \leq O(1)p^{-1}b.
\end{aligned}$$

Since $b/p \to 0$, $\operatorname{var}(\mathcal{U}_3) = o(1)$.

Given $\operatorname{var}(\mathcal{U}_2) = o(1)$ and $\operatorname{var}(\mathcal{U}_3) = o(1)$ above, next we focus on $\mathcal{U}_1$. By the α-mixing assumption in Condition 3.3.3, and following the similar arguments in (Ibragimov and Linnik, 1971, page 338), we have for properly chosen r and b_2,

$$\left| \mathrm{E}\left\{\exp(it\mathcal{U}_1)\right\} - \prod_{k=1}^{r} \mathrm{E}\left[\exp\left\{it\sigma^{-1}(a)A_{k,1}(a)\right\}\right] \right| \leq 16r\alpha(b_2) \to 0.$$

This suggests there exist independent random variables $\{\xi_k : k = 1, \cdots, r\}$ such that ξ_k and $A_{k,1}(a)$ are identically distributed and $\mathcal{U}_1$ has the same asymptotic distribution as $\sigma^{-1}(a)\sum_{k=1}^{r}\xi_k$. To prove the asymptotic normality of $\sigma^{-1}(a)\mathcal{U}_1$, now it remains to show that central limit theorem holds for $\sigma^{-1}(a)\sum_{k=1}^{r}\xi_k$. Then we check the Lyapunov condition, i.e., check that the moments of ξ_k satisfy

$$s_r^{-4}\sum_{k=1}^{r} \mathrm{E}\left\{\sigma^{-1}(a)|\xi_k|\right\}^4 \to 0, \tag{B.173}$$

where we define $s_r^2 = \sum_{k=1}^{r}\mathrm{var}\{\sigma^{-1}(a)\xi_k\}$. By Lemma B.5.8, for even $\epsilon > 0$,

$$\mathrm{M}_{4+\epsilon} := \max_{1\leq j\leq p}\left\{\left\|n^{a/2}\left\{\mathcal{U}^{(j)}(a)\right\}\right\|_{4+\epsilon}\right\} < \infty. \tag{B.174}$$

Then by the moment bounds in (Kim, 1994, Theorem 1), and the α-mixing assumption in Condition 3.3.3, for $g(2,\epsilon) = \epsilon/(4+\epsilon)$,

$$\mathrm{E}\left(\left[\sum_{j=1}^{b_1} n^{a/2}\left\{\mathcal{U}^{(j)}(a)\right\}\right]^4\right) \leq Cb_1^2\Big\{C + M_{4+\epsilon}^4\sum_{j=1}^{b_1} j^{2-1}\alpha(j)^{g(2,\epsilon)}\Big\}$$

As $\delta \in (0,1)$ and $0 < g(2,\epsilon) < 1$,

$$\sum_{j=1}^{\infty} j\alpha(j)^{g(2,\epsilon)} \leqslant C\sum_{j=1}^{\infty} j \times (\delta^{g(2,\epsilon)})^j < \infty.$$

It follows that

$$
\begin{aligned}
\mathrm{E}\left\{\sigma^{-1}(a)A_{1,1}(a)\right\}^4 &= \sigma^{-4}(a)n^{-2a}\mathrm{E}\left[\sum_{j=1}^{b_1} n^{a/2}\left\{\mathcal{U}^{(j)}(a)\right\}\right]^4 \\
&\leq O(1)p^{-2}n^{2a}n^{-2a} \times b_1^2\left\{C + M_{4+\epsilon}^4\sum_{j=1}^{b_1} j^{2-1}\alpha(j)^{g(2,\epsilon)}\right\} \\
&= O(1)p^{-2} \times b_1^2.
\end{aligned}
$$

Similarly, for other $k > 1$, $\mathrm{E}\left\{\sigma^{-1}(a)A_{k,1}(a)\right\}^4$ have the same bound. Thus,

$$
\sum_{k=1}^{r} \sigma^{-4}(a)E|\xi_k|^4 = O(1)rp^{-2}b_1^2. \tag{B.175}
$$

In addition,

$$
\begin{aligned}
\mathrm{var}\{\sigma^{-1}(a)\xi_k\} &= \sigma^{-2}(a)\mathrm{var}\left\{\sum_{i=1}^{b_1}\mathcal{U}^{((k-1)b+i)}(a)\right\} \\
&= \sigma^{-2}(a)\sum_{1\leq i_1,i_2\leq b_1} \mathrm{cov}\left\{\mathcal{U}^{((k-1)b+i_1)}(a), \mathcal{U}^{((k-1)b+i_2)}(a)\right\} \\
&= \sigma^{-2}(a)\sum_{1\leq i_1,i_2\leq b_1} (\mathrm{B.169}).
\end{aligned}
$$

By Condition 3.3.3 and $rb_1/p \to 1$, we have

$$
\begin{aligned}
s_r^4 &= \left[\sum_{j=1}^{r} \mathrm{var}\left\{\xi_j/\sigma(a)\right\}\right]^2 \\
&= \Theta(1)p^{-2}n^{2a}(r \times b_1 n^{-a})^2 = \Theta(1)p^{-2}r^2b_1^2.
\end{aligned} \tag{B.176}
$$

Combine (B.175) and (B.176), (B.173) is proved as $r \to \infty$.

In summary, for any finite integer a, we prove the asymptotic normality of $\mathcal{U}(a)/\sigma(a)$. For any linear combination of U-statistics $Z_n := \sum_{r=1}^{m} t_r\mathcal{U}(a_r)/\sigma(a_r)$, we can similarly decompose Z_n into three parts and apply the analysis above. The similar conclusion holds for finite m and the asymptotic joint normality is obtained by the Cramér-Wold

Theorem.

Lemma B.5.8. *For $\forall$ finite even $\omega > 0$ any $\forall$ finite integer $a > 0$,*

$$\max_{1\le j\le p} \mathrm{E}\left\{n^{a/2}\mathcal{U}^{(j)}(a)\right\}^{\omega} < \infty.$$

Proof. Recall the definition of $\mathcal{U}^{(j)}(a)$ in (B.167). For positive even ω,

$$\begin{aligned}
&\mathrm{E}[\{\mathcal{U}^{(j)}(a)\}^{\omega}] \\
= &\sum_{l=1}^{\omega} \sum_{\substack{0\le c_l\le a,\\ \mathbf{k}^{(l)}\in\mathcal{P}(n_x,c_l),\\ \mathbf{s}^{(l)}\in\mathcal{P}(n_y,a-c_l)}} G(c_l)\mathrm{E}\left(\prod_{l=1}^{\omega}\prod_{t_l=1}^{c_l} x_{k_{t_l}^{(l)},j}\right)\mathrm{E}\left(\prod_{l=1}^{\omega}\prod_{m_l=1}^{a-c_l} y_{s_{m_l}^{(l)},j}\right).
\end{aligned} \tag{B.177}$$

Define the index tuple $(\mathbf{k}^{(1)},\dots,\mathbf{k}^{(\omega)}) = (k_1^{(1)},\dots,k_{c_1}^{(1)},\dots,k_1^{(\omega)},\dots,k_{c_\omega}^{(\omega)})$. When $|\{(\mathbf{k}^{(1)},\dots,\mathbf{k}^{(\omega)})\}| > \sum_{l=1}^{\omega} c_l/2$, it means that one of the index appears only once. Suppose index $i \in \{(\mathbf{k}^{(1)},\dots,\mathbf{k}^{(\omega)})\}$ only appears once, then under H_0,

$$\mathrm{E}\left(\prod_{l=1}^{\omega}\prod_{t_l=1}^{c_l} x_{k_{t_l}^{(l)},j}\right) = \mathrm{E}(x_{i,j}) \times \mathrm{E}(\text{other terms}) = 0. \tag{B.178}$$

Thus (B.178) $\neq 0$ only when $|\{(\mathbf{k}^{(1)},\dots,\mathbf{k}^{(\omega)})\}| \le \sum_{l=1}^{\omega} c_l/2$. By the boundedness of moments in Condition 3.3.3,

$$\max_{1\le j\le p} \sum_{0\le c_l\le a,\, \mathbf{k}^{(l)}\in\mathcal{P}(n_x,c_l)} \mathrm{E}\left(\prod_{l=1}^{\omega}\prod_{t_l=1}^{c_l} x_{k_{t_l}^{(l)},j}\right) = O\left(n_x^{\sum_{l=1}^{\omega} c_l/2}\right).$$

Similarly, we have

$$\max_{1\le j\le p} \sum_{0\le c_l\le a,\, \mathbf{s}^{(l)}\in\mathcal{P}(n_y,a-c_l)} \mathrm{E}\left(\prod_{l=1}^{\omega}\prod_{m_l=1}^{a-c_l} y_{s_{m_l}^{(l)},j}\right) = O\left(n_y^{\sum_{l=1}^{\omega} (a-c_l)/2}\right).$$

As $G(a,c) = \Theta(n_x^{-c}n_y^{-(a-c)})$, by (B.177), $\max_{1\le j\le p}\mathrm{E}[\{n^{a/2}\mathcal{U}^{(j)}(a)\}^{\omega}] < \infty$. □

B.5.20 Proof of Lemma B.2.5

Recall $\mathcal{U}^{(j)}(a)$ defined in (B.167). Similarly to $\tilde{\mathcal{U}}_c(a)$, we define $\tilde{\mathcal{U}}_c^{(j)}(a)$ as the sequence of random variables on the conditional probability measure $\tilde{P}$, given the event $n_x n_y \mathcal{U}(\infty)/(n_x+n_y) - \tau_p \le u$ such that

$$\begin{aligned}
&\tilde{P}\Big\{\tilde{\mathcal{U}}_c^{(j)}(a) \le u_j : 1 \le j \le p\Big\} \\
&= P\Big\{\mathcal{U}^{(j)}(a) \le u_j : 1 \le j \le p \;\Big|\; \frac{n_x n_y}{n_x+n_y}\mathcal{U}(\infty) \le \tau_p + u\Big\}.
\end{aligned}$$

Then $\sigma^{-1}(a)\tilde{\mathcal{U}}_c(a) = \sigma^{-1}(a)\sum_{j=1}^{p}\tilde{\mathcal{U}}_c^{(j)}(a)$, and we prove the asymptotic normality of $\sigma^{-1}(a)\tilde{\mathcal{U}}_c(a)$ similarly to Section B.5.19. In particular, we partition the sequence $\{\sigma^{-1}(a) \times \tilde{\mathcal{U}}_c^{(j)}(a) : 1 \le j \le p\}$ into r blocks, where each block contains b variables such that $rb \le p < (r+1)b$. For each $1 \le k \le r$, we further partition the kth block into two sub-blocks such that a larger one $\tilde{A}_{k,1}$ contains the first b_1 variables and a smaller one $\tilde{A}_{k,2}$ contains the last $b_2 = b - b_1$ variables. Similarly, for $1 \le k \le r$, we write

$$\tilde{A}_{k,1}(a) = \sum_{i=1}^{b_1}\tilde{\mathcal{U}}_c^{(k-1)b+i}(a), \quad \tilde{A}_{k,2}(a) = \sum_{i=1}^{b_2}\tilde{\mathcal{U}}_c^{(k-1)b+b_1+i}(a).$$

Correspondingly, define $\tilde{\mathcal{U}}_1 = \sigma^{-1}(a)\sum_{k=1}^{r}\tilde{A}_{k,1}(a)$, $\tilde{\mathcal{U}}_2 = \sigma^{-1}(a)\sum_{k=1}^{r}\tilde{A}_{k,2}(a)$ and $\tilde{\mathcal{U}}_3 = \sigma^{-1}(a)\sum_{j=rb+1}^{p}\tilde{\mathcal{U}}_c^{(j)}(a)$. Then we have the decomposition: $\sigma^{-1}(a) \times \tilde{\mathcal{U}}_c(a) = \tilde{\mathcal{U}}_1 + \tilde{\mathcal{U}}_2 + \tilde{\mathcal{U}}_3$. To show that $\sigma^{-1}(a) \times \tilde{\mathcal{U}}_c(a)$ satisfies the central limit theorem, we first show that $\tilde{\mathrm{E}}(\tilde{\mathcal{U}}_2^2) = o(1)$ and $\tilde{\mathrm{E}}(\tilde{\mathcal{U}}_3^2) = o(1)$.

$$
\begin{aligned}
\tilde{\mathrm{E}}(\tilde{\mathcal{U}}_2^2) &= \sigma^{-2}(a)\tilde{\mathrm{E}}\Big\{\Big(\sum_{k=1}^{r} \tilde{A}_{k,2}(a)\Big)^2\Big\} \\
&\leq \sigma^{-2}(a)\Bigg(\sum_{1\leq k_1,k_2\leq r} \Big[\tilde{\mathrm{E}}\Big\{\tilde{A}^2_{k_1,2}(a)\Big\}\Big]^{1/2}\Big[\tilde{\mathrm{E}}\Big\{\tilde{A}^2_{k_2,2}(a)\Big\}\Big]^{1/2}\Bigg) \\
&\leq \sigma^{-2}(a)\Big[P\Big\{\frac{n_x n_y}{n_x+n_y}\mathcal{U}(\infty) < \tau_p\Big\}\Big]^{-1} \\
&\quad\times\Bigg(\sum_{1\leq k_1,k_2\leq r} \Big[\mathrm{E}\Big\{A^2_{k_1,2}(a)\Big\}\Big]^{1/2}\Big[\mathrm{E}\Big\{A^2_{k_2,2}(a)\Big\}\Big]^{1/2}\Bigg),
\end{aligned}
$$

where in the last inequality we use the fact that

$$
\begin{aligned}
\tilde{\mathrm{E}}\Big\{\tilde{A}^2_{k,2}(a)\Big\} &= \frac{\mathrm{E}\{A^2_{k,2}(a)\mathbf{1}_{\{n_x n_y \mathcal{U}(\infty)/(n_x+n_y)<\tau_p+u\}}\}}{P\{n_x n_y \mathcal{U}(\infty)/(n_x+n_y) < \tau_p + u\}} \\
&\leq \frac{\mathrm{E}\{A^2_{k,2}(a)\}}{P\{n_x n_y \mathcal{U}(\infty)/(n_x+n_y) < \tau_p + u\}}.
\end{aligned}
$$

The upper bound above converges to 0 under the α-mixing condition by choosing proper convergence rate b_2; see Eq. (18.4.8) of Ibragimov and Linnik (1971). Similarly, we can also show $\tilde{\mathrm{E}}(\tilde{\mathcal{U}}_3^2) = o(1)$. It remains to examine the $\tilde{\mathcal{U}}_1$. Define $\alpha(s)$ as the mixing coefficient of $\{(x_{1,j},\ldots,x_{n_x,j},y_{1,j},\ldots,y_{n_y,j} : j = 1,\ldots,p)\}$ and define $\tilde{\alpha}(s)$ as the corresponding mixing coefficient on the conditional probability measure. Following a similar argument to that in (Hsing, 1995, Lemma 2.2), we have

$$
\tilde{\alpha}(d) \leq 4\frac{\max_{1\leq h\leq p-d} P\{U^0_{h,d}(\infty) > \tau_p + u\} + \alpha(d)}{[P\{n_x n_y \mathcal{U}(\infty)/(n_x+n_y) < \tau_p + u\}]^3},
$$

where $U^0_{h,d}(\infty) = \max_{h\leq j\leq h+d} U^{(j)}(\infty)$, $U^{(j)}(\infty) = \sigma^{-1}_{j,j} \times (\bar{x}_j - \bar{y}_j)^2 \times n_x n_y/(n_x + n_y)$, and recall $\tau_p = 2\log p - \log\log p$. Since $x_{i,j}$ and $y_{i,j}$ are sub-gaussian random variables by Condition 3.3.3 (Vershynin, 2018, Proposition 2.5.2), we know $\sigma^{-1/2}_{j,j} \times (\bar{x}_j - \bar{y}_j) \times \sqrt{n_x n_y}/\sqrt{n_x + n_y}$ is a sub-gaussian variable with variance 1. Therefore, $\max_{1\leq h\leq p-d} P\{U^0_{h,d}(\infty) > \tau_p + u\} \leq d\max_{1\leq j\leq p} P\{U^{(j)}(\infty) > \tau_p + u\} \leq$

$Cd\exp\{-(\tau_p+u)/2\} \le Cdp^{-1}\sqrt{\log p}$. Then similarly to Page 338 in Ibragimov and Linnik (1971), we have

$$
\begin{aligned}
& \left|\tilde{\mathrm{E}}\left\{\exp(it\tilde{\mathcal{U}}_1)\right\} - \prod_{k=1}^{r}\tilde{\mathrm{E}}\left[\exp\left\{it\sigma^{-1}(a)\tilde{A}_{k,1}(a)\right\}\right]\right| \\
\le\ & 16r\tilde{\alpha}(b_2) \\
\le\ & 64r\frac{\max_{1\le h\le p-b_2} P\{U^0_{h,b_2}(\infty) > \tau_p+u\} + \alpha(b_2)}{[P\{n_xn_y\mathcal{U}(\infty)/(n_x+n_y) < \tau_p+u\}]^3},
\end{aligned}
$$

which converges to 0 for properly chosen r and b_2 such that $rb_2\sqrt{\log p}/p \to 0$. Thus there exist independent $\{\tilde{\xi}_k : k=1,\ldots,r\}$ such that $\tilde{\xi}_k$ and $\tilde{A}_{k1}(a)$ are identically distributed on probability measure $\tilde{P}$. Similarly to (Hsing, 1995, Lemma 2.4, Lemma 2.5), we have $\tilde{\mathrm{E}}\{\sigma^{-1}(a)\sum_{k=1}^{r}\tilde{\xi}_k\} \to 0$ and $\tilde{\mathrm{E}}[\{\sigma^{-1}(a)\sum_{k=1}^{r}\tilde{\xi}_k\}^2] \to 1$. To show the asymptotic normality on the conditional probability measure, it remains to check the Lyapunov condition that

$$
\sum_{k=1}^{r}\tilde{\mathrm{E}}\left\{\sigma^{-1}(a)|\tilde{\xi}_k|\right\}^4 \le \sigma^{-4}(a)\frac{\sum_{k=1}^{r}\mathrm{E}(\xi_k^4)}{P\{n_xn_y\mathcal{U}(\infty)/(n_x+n_y) < \tau_p+u\}} \to 0,
$$

where ξ_k are define same as in Appendix Section B.5.19, and the convergence result follows from (B.173). This implies the asymptotic normality of conditional distribution given $\{n_xn_y\mathcal{U}(\infty)/(n_x+n_y) < \tau_p+u\}$. Thus we obtain the asymptotic independence between $\mathcal{U}(a)/\sigma(a)$ and $\mathcal{U}(\infty)$.

B.5.21 Proof of Lemma B.2.6

Recall the definitions in (B.22). $T_{a,2}$ is the summation over j indexes in the set $\{k_0,\ldots,p\}$ such that $\mu_j = \nu_j = 0$. Then $\mathrm{E}(T_{a,2}) = 0$. Following the argument in Section B.5.17, we obtain

$$
\mathrm{var}(T_{a,2}) \simeq \sum_{k_0+1\le j_1,j_2\le p} a!\Big(\frac{\sigma_{x,j_1,j_2}}{n_x} + \frac{\sigma_{y,j_1,j_2}}{n_y}\Big)^a.
$$

Let $\mathcal{V}_{a,j_1,j_2} = \{\sigma_{x,j_1,j_2}/\gamma + \sigma_{y,j_1,j_2}/(1-\gamma)\}^a$. By the mixing assumption in Condition 3.3.3 and Lemma B.5.1, we know there exist some constants C and $\tilde{\delta}$ such that $|\mathcal{V}_{a,j_1,j_2}| \leq C\tilde{\delta}^{|j_1-j_2|}$. Note that

$$\begin{aligned}
&\Big|\sum_{1\leq j_1,j_2\leq p}\mathcal{V}_{a,j_1,j_2} - \sum_{k_0+1\leq j_1,j_2\leq p}\mathcal{V}_{a,j_1,j_2}\Big| \\
&= \Big|\Big(\sum_{1\leq j_1,j_2\leq k_0} + \sum_{1\leq j_1\leq k_0,\, k_0+1\leq j_2\leq p} + \sum_{1\leq j_2\leq k_0,\, k_0+1\leq j_1\leq p}\Big)\mathcal{V}_{a,j_1,j_2}\Big| \\
&\leq C\Big(\sum_{1\leq j_1,j_2\leq k_0} + \sum_{1\leq j_1\leq k_0,\, k_0+1\leq j_2\leq p} + \sum_{1\leq j_2\leq k_0,\, k_0+1\leq j_1\leq p}\Big)\tilde{\delta}^{|j_1-j_2|} = O(k_0).
\end{aligned}$$

Since $k_0 = o(p)$ and Condition 3.3.3 assumes that $\sum_{1\leq j_1,j_2\leq p}\mathcal{V}_{a,j_1,j_2} = \Theta(p)$, then $\sum_{k_0+1\leq j_1,j_2\leq p}\mathcal{V}_{a,j_1,j_2} = \Theta(p)$. It follows that $\mathrm{var}(T_{a,2}) = \Theta(p^2 n^{-a})$.

It remains to prove $\mathrm{var}(T_{a,1}) = o(pn^{-a})$. Note that $\mathrm{var}(T_{a,1}) = \mathrm{E}(T_{a,1}^2) - \{\mathrm{E}(T_{a,1})\}^2$, and $\mathrm{E}(T_{a,1}) = k_0\rho^a$. Following the definition in (B.22),

$$\mathrm{E}(T_{a,1}^2) = \sum_{1\leq j_1,j_2\leq k_0}\ \sum_{\substack{0\leq c\leq a,\\ \mathbf{k}\in\mathcal{P}(n_x,c),\\ \mathbf{s}\in\mathcal{P}(n_y,a-c)}}\ \sum_{\substack{0\leq \tilde{c}\leq a,\\ \tilde{\mathbf{k}}\in\mathcal{P}(n_x,\tilde{c}),\\ \tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-\tilde{c})}} G(a,c)G(a,\tilde{c})Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}),$$

where similarly to Section B.5.17,

$$Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}) = \mathrm{E}\Big(\prod_{t=1}^{c} x_{k_t,j_1}\prod_{\tilde{t}=1}^{\tilde{c}} x_{\tilde{k}_{\tilde{t}},j_2}\Big)\mathrm{E}\Big(\prod_{m=1}^{a-c} y_{s_m,j_1}\prod_{\tilde{m}=1}^{a-\tilde{c}} y_{\tilde{s}_{\tilde{m}},j_2}\Big).$$

Since $\mathrm{E}(\mathbf{y}) = \boldsymbol{\nu} = \mathbf{0}$, if $\{\mathbf{s}\} \neq \{\tilde{\mathbf{s}}\}$, $Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j}) = 0$. If $\{\mathbf{s}\} = \{\tilde{\mathbf{s}}\}$, it induces $c = \tilde{c}$. When $\{\mathbf{s}\} = \{\tilde{\mathbf{s}}\}$, let $b = |\{\mathbf{k}\}\cap\{\tilde{\mathbf{k}}\}|$, then $0 \leq b \leq c$, $\mathrm{E}\{Q(\mathbf{k},\mathbf{s},\tilde{\mathbf{k}},\tilde{\mathbf{s}},\mathbf{j})\} = \mu_{j_1}^{c-b}\mu_{j_2}^{c-b}\varphi_{j_1,j_2}^{b}\sigma_{j_1,j_2}^{a-c} = \rho^{2(c-b)}\varphi_{j_1,j_2}^{b}\sigma_{j_1,j_2}^{a-c}$, and

$$\mathrm{E}(T_{a,1}^2) = \sum_{1\leq j_1,j_2\leq k_0}\ \sum_{\substack{0\leq c\leq a,\\ \mathbf{k},\tilde{\mathbf{k}}\in\mathcal{P}(n_x,c);\\ \mathbf{s},\tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-c)}} G^2(a,c)\times\rho^{2(c-b)}\varphi_{j_1,j_2}^{b}\sigma_{j_1,j_2}^{a-c}\times\mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\}\}}.$$

We next decompose $\mathrm{E}(T_{1,a}^2) = G_{t,1,a,1} + G_{t,1,a,2} + G_{t,1,a,3}$, where

$$G_{t,1,a,1} = \sum_{1 \le j_1, j_2 \le k_0} \sum_{\substack{0 \le c \le a, \\ \mathbf{k}, \tilde{\mathbf{k}} \in \mathcal{P}(n_x, c); \\ \mathbf{s}, \tilde{\mathbf{s}} \in \mathcal{P}(n_y, a-c)}} G^2(a,c) \rho^{2(c-b)} \varphi_{j_1,j_2}^b \sigma_{j_1,j_2}^{a-c} \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\}, c=a, b=0\}},$$

$$G_{t,1,a,2} = \sum_{1 \le j_1, j_2 \le k_0} \sum_{\substack{0 \le c \le a, \\ \mathbf{k}, \tilde{\mathbf{k}} \in \mathcal{P}(n_x, c); \\ \mathbf{s}, \tilde{\mathbf{s}} \in \mathcal{P}(n_y, a-c)}} G^2(a,c) \rho^{2(c-b)} \varphi_{j_1,j_2}^b \sigma_{j_1,j_2}^{a-c} \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\}, c \le a-1, b=0\}},$$

and

$$G_{t,1,a,3} = \sum_{1 \le j_1, j_2 \le k_0} \sum_{\substack{0 \le c \le a, \\ \mathbf{k}, \tilde{\mathbf{k}} \in \mathcal{P}(n_x, c); \\ \mathbf{s}, \tilde{\mathbf{s}} \in \mathcal{P}(n_y, a-c)}} G^2(a,c) \rho^{2(c-b)} \varphi_{j_1,j_2}^b \sigma_{j_1,j_2}^{a-c} \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\}, 1 \le b \le c\}}.$$

Note that $|\mathrm{var}(T_{a,1})| \le |G_{t,1,a,1} - \{\mathrm{E}(T_{a,1})\}^2| + |G_{t,1,a,2}| + |G_{t,1,a,3}|$. To prove $\mathrm{var}(T_{a,1}) = o(pn^{-a})$, we will next show $|G_{t,1,a,1} - \{\mathrm{E}(T_{a,1})\}^2|$, $|G_{t,1,a,2}|$ and $|G_{t,1,a,3}|$ are $o(pn^{-a})$ respectively.

First, as $\sum_{\mathbf{k}, \tilde{\mathbf{k}} \in \mathcal{P}(n_x, a); \mathbf{s}, \tilde{\mathbf{s}} \in \mathcal{P}(n_y, a-c)} \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\}, c=a, b=0\}} = P_{2a}^{n_x}$ and $G(a,a) = (P_a^{n_x})^{-1}$,

$$G_{t,1,a,1} = \sum_{1 \le j_1, j_2 \le k_0} \sum_{\substack{0 \le c \le a, \\ \mathbf{k}, \tilde{\mathbf{k}} \in \mathcal{P}(n_x, c); \\ \mathbf{s}, \tilde{\mathbf{s}} \in \mathcal{P}(n_y, a-c)}} G^2(a,c) \rho^{2a} \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\}, c=a, b=0\}} = \frac{P_{2a}^{n_x}}{(P_a^{n_x})^2} k_0^2 \rho^{2a}.$$

Then $|G_{t,1,a,1} - \{\mathrm{E}(T_{a,1})\}^2| = o(1) k_0^2 n^{-2a} n^{2a} \rho^{2a} = o(pn^{-a})$, where we use $\mathrm{E}(T_{a,1}) = k_0 \rho^a$. In addition, as $\sum_{\mathbf{k}, \tilde{\mathbf{k}} \in \mathcal{P}(n_x, c); \mathbf{s}, \tilde{\mathbf{s}} \in \mathcal{P}(n_y, a-c)} \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\}, c \le a-1, b=0\}} = O(n^{2c+a-c})$ and $G(a,c) = \Theta(n^{-a})$, we have $|G_{t,1,a,2}| \le C \sum_{1 \le j_1, j_2 \le k_0} \sum_{c=0}^{a-1} n^{-(a-c)} \rho^{2c} \sigma_{j_1,j_2}^{a-c}$. Since we have $\sum_{1 \le j_1, j_2 \le k_0} \sigma_{j_1,j_2} = O(k_0)$ by Condition 3.3.3 and Lemma B.5.1, we further know $|G_{t,1,a,2}| = \sum_{c=0}^{a-1} O(k_0 \rho^{2c} n^{-(a-c)})$. As $\rho = O(k_0^{-1/a} p^{1/(2a)} n^{-1/2})$ and $k_0 = o(p)$, we obtain $|G_{t,1,a,2}| = o(pn^{-a})$. Moreover, as $G(a,c) = \Theta(n^{-a})$, $\varphi_{j_1,j_2} = \rho^2 + \sigma_{j_1,j_2}$, and

$\sum_{\mathbf{k},\tilde{\mathbf{k}}\in\mathcal{P}(n_x,c);\mathbf{s},\tilde{\mathbf{s}}\in\mathcal{P}(n_y,a-c)} \mathbf{1}_{\{\{\mathbf{s}\}=\{\tilde{\mathbf{s}}\},b\geq 1\}} = O(n^{2c-b+a-c})$,

$$|G_{t,1,a,3}| \leq C \sum_{\substack{0\leq c\leq a,\\ 1\leq b\leq c}} \sum_{1\leq j_1,j_2\leq k_0} n^{-(b+a-c)}\rho^{2(c-b)}(\sigma_{j_1,j_2}+\rho^2)^b_{j_1,j_2}\sigma^{a-c}_{j_1,j_2}.$$

For given c and b, the maximum order of $\sum_{1\leq j_1,j_2\leq k_0} n^{-(b+a-c)}\rho^{2(c-b)}(\sigma_{j_1,j_2}+\rho^2)^b_{j_1,j_2}\sigma^{a-c}_{j_1,j_2}$ is bounded by the following two quantities:

$$\sum_{1\leq j_1,j_2\leq k_0} Cn^{-(b+a-c)}\rho^{2c}\sigma^{a-c}_{j_1,j_2}, \tag{B.179}$$

$$\sum_{1\leq j_1,j_2\leq k_0} Cn^{-(b+a-c)}\sigma^{b+a-c}_{j_1,j_2}\rho^{2(c-b)}. \tag{B.180}$$

For (B.179), when $c = a$, (B.179) $= O(k_0^2 n^{-b}\rho^{2a}) = o(pn^{-a})$. When $c \leq a-1$, since $\sum_{1\leq j_1,j_2\leq k_0}\sigma_{j_1,j_2} = O(k_0)$ by Condition 3.3.3 and Lemma B.5.1, then (B.179) $= O(k_0 n^{-(b+a-c)}\rho^{2c}) = o(pn^{-a})$. For (B.180), as $b\geq 1$, $b+a-c\geq 1$. Then similarly by Condition 3.3.3 and Lemma B.5.1, (B.180) $= O(k_0 n^{-(b+a-c)}\rho^{2(c-b)}) = o(pn^{-a})$.

In summary, we obtain $\text{var}(T_{a,1}) = o(pn^{-a}) = o(1)\text{var}(T_{a,2})$. Then

$$\text{var}\{\mathcal{U}(a)\} \simeq \text{var}(T_{a,2}) \simeq \sum_{k_0+1\leq j_1,j_2\leq p} a!\Big(\frac{\sigma_{x,j_1,j_2}}{n_x} + \frac{\sigma_{y,j_1,j_2}}{n_y}\Big)^a.$$

By Markov's inequality, $\{T_{a,1} - \text{E}(T_{a,1})\}/\sigma(a) \xrightarrow{P} 0$.

B.5.22 Proof of Lemma B.2.7

Note that

$$\{\sigma(a)\sigma(b)\}^{-1}\text{cov}\{\mathcal{U}(a),\mathcal{U}(b)\} = \{\sigma(a)\sigma(b)\}^{-1} \times \sum_{1\leq l_1,l_2\leq 2} \text{cov}(T_{a,l_1}, T_{b,l_2}).$$

Lemma B.2.6 suggests that $\text{var}(T_{a,1}) = o(1)\sigma^2(a)$. By the Cauchy-Schwarz inequality, $\{\sigma(a)\sigma(b)\}^{-1}\text{cov}\{\mathcal{U}(a),\mathcal{U}(b)\} = \{\sigma(a)\sigma(b)\}^{-1}\text{cov}(T_{a,2}, T_{b,2}) + o(1)$. To finish the proof, it suffices to show $\mathrm{cov}(T_{a,2}, T_{b,2}) = 0$. Note that $T_{a,2}$ and $T_{b,2}$ are summation over j indexes in the set $\{k_0, \ldots, p\}$ such that $\mu_j = \nu_j = 0$. Then the proof in Section B.5.18 applies similarly and we have $\mathrm{cov}(T_{a,2}, T_{b,2}) = 0$.

B.5.23 Proof of Lemma B.3.1

In the following, we will first derive the form of $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ and then prove that $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$.

As we assume $\mathrm{E}(\mathbf{x}) = \mathrm{E}(\mathbf{y}) = \mathbf{0}$, $\mathrm{cov}(x_{1,j_1}, x_{1,j_2}) = \mathrm{E}(x_{1,j_1}x_{1,j_2})$ and $\mathrm{cov}(y_{1,j_1}, y_{1,j_2}) = \mathrm{E}(y_{1,j_1}y_{1,j_2})$. It follows that $\mathrm{E}\{\tilde{\mathcal{U}}(a)\} = 0$ and $\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \mathrm{E}\{\tilde{\mathcal{U}}^2(a)\}$. By definition,

$$\tilde{\mathcal{U}}(a) = (P_a^{n_x} P_a^{n_y})^{-1} \sum_{1 \le j_1, j_2 \le p} \sum_{\substack{\mathbf{i} \in \mathcal{P}(n_x, a); \\ \mathbf{w} \in \mathcal{P}(n_y, a)}} \mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i}, \mathbf{w}, j_1, j_2),$$

where we define $\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i}, \mathbf{w}, j_1, j_2) = \prod_{t=1}^{a}(x_{i_t,j_1}x_{i_t,j_2} - y_{w_t,j_1}y_{w_t,j_2})$. Then

$$\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \frac{1}{(P_a^{n_x} P_a^{n_y})^2} \sum_{\substack{1 \le j_1, j_2, j_3, j_4 \le p; \\ \mathbf{i}, \tilde{\mathbf{i}} \in \mathcal{P}(n_x, a); \\ \mathbf{w}, \tilde{\mathbf{w}} \in \mathcal{P}(n_y, a)}} \mathrm{E}\Big\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i}, \mathbf{w}, j_1, j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}}, \tilde{\mathbf{w}}, j_3, j_4)\Big\}.$$

Under H_0, $\Sigma_x = \Sigma_y = \Sigma = (\sigma_{j_1,j_2})_{p\times p}$, then $\mathrm{E}(x_{1,j_1}x_{1,j_2} - \sigma_{j_1,j_2}) = 0$ and $\mathrm{E}(y_{1,j_1}y_{1,j_2} - \sigma_{j_1,j_2}) = 0$. If $|\{\mathbf{i}\} \cap \{\tilde{\mathbf{i}}\}| + |\{\mathbf{w}\} \cap \{\tilde{\mathbf{w}}\}| < a$, it means that the common indexes between $(\mathbf{i}, \mathbf{w})$ and $(\tilde{\mathbf{i}}, \tilde{\mathbf{w}})$ is smaller than a, then we know $\mathrm{E}\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i}, \mathbf{w}, j_1, j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}}, \tilde{\mathbf{w}}, j_3, j_4)\} = 0$. If $|\{\mathbf{i}\} \cap \{\tilde{\mathbf{i}}\}| + |\{\mathbf{w}\} \cap \{\tilde{\mathbf{w}}\}| \ge a$, we know $\mathrm{E}\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i}, \mathbf{w}, j_1, j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}}, \tilde{\mathbf{w}}, j_3, j_4)\}$ is a linear combination of $(\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m}$, where $a - |\{\mathbf{w}\} \cap \{\tilde{\mathbf{w}}\}| \le m \le |\{\mathbf{i}\} \cap \{\tilde{\mathbf{i}}\}|$ and

$$\begin{aligned}
\mathbf{X}_{j_1,j_2,j_3,j_4} &= \mathrm{E}\{(x_{1,j_1}x_{1,j_2} - \sigma_{j_1,j_2})(x_{1,j_3}x_{1,j_4} - \sigma_{j_3,j_4})\}, \\
\mathbf{Y}_{j_1,j_2,j_3,j_4} &= \mathrm{E}\{(y_{1,j_1}y_{1,j_2} - \sigma_{j_1,j_2})(y_{1,j_3}y_{1,j_4} - \sigma_{j_3,j_4})\}.
\end{aligned}$$

And if $|\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}|+|\{\mathbf{w}\}\cap\{\tilde{\mathbf{w}}\}|=t_0$,

$$\sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a);\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} \mathbf{1}_{\{|\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}|+|\{\mathbf{w}\}\cap\{\tilde{\mathbf{w}}\}|=t_0\}} = O(n^{4a-t_0}),$$

which achieves the largest order at $t_0=a$ when $t_0\geq a$. Therefore,

$$\begin{aligned}\mathrm{var}\{\tilde{\mathcal{U}}(a)\} \;\simeq\; & \frac{1}{(P_a^{n_x}P_a^{n_y})^2}\sum_{\substack{1\leq j_1,j_2,j_3,j_4\leq p;\\ \mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a);\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} \mathbf{1}_{\{|\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}|+|\{\mathbf{w}\}\cap\{\tilde{\mathbf{w}}\}|=a\}} \\ & \times \mathrm{E}\Big\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},j_3,j_4)\Big\}.\end{aligned}$$

It follows that

$$\begin{aligned}\mathrm{var}\{\tilde{\mathcal{U}}(a)\} \;\simeq\; & \sum_{1\leq j_1,j_2,j_3,j_4\leq p}\sum_{m=0}^{a}\frac{P_{2a-m}^{n_x}P_{a+m}^{n_y}}{(P_a^{n_x}P_a^{n_y})^2}\binom{a}{m}^2\binom{a-m}{a-m}^2 \\ & \times m!(a-m)!(\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m},\end{aligned} \tag{B.181}$$

and then (B.181) $\simeq \sum_{1\leq j_1,j_2,j_3,j_4\leq p} a!(\mathbf{X}_{j_1,j_2,j_3,j_4}/n_x+\mathbf{Y}_{j_1,j_2,j_3,j_4}/n_y)^a$.

We next prove $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}=o(1)\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$ under Conditions 3.4.1 and 3.4.1* in the following Sections B.5.23.1 and B.5.23.2 respectively.

B.5.23.1 Proof under Condition 3.4.1

To prove $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}=o(1)\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$ under Condition 3.4.1, we will first show $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}=\Theta(p^2n^{-a})$. Note that $P_{2a-m}^{n_x}P_{a+m}^{n_y}/(P_a^{n_x}P_a^{n_y})^2\simeq Cn^a$. By (B.181), it remains to show that for any $m\in\{0,1,\ldots,a\}$,

$$\sum_{1\leq j_1,j_2,j_3,j_4\leq p}(\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m}=\Theta(p^2). \tag{B.182}$$

We next prove (B.182) by discussing different cases of $\{j_1, j_2, j_3, j_4\}$, and using $K_0 = -(2+\epsilon)(8+2\mu)(\log p)/(\epsilon \log \delta)$ similarly to (B.77), where ϵ and μ are positive constants and $\delta = \max\{\delta_x, \delta_y\}$ from Condition 3.4.1.

Case 1: If $|j_1 - j_2| \le K_0$ and $|j_3 - j_4| \le K_0$, we define a distance $\kappa_d = \min\{|j_1 - j_3|, |j_1 - j_4|, |j_2 - j_3|, |j_2 - j_4|\}$, and discuss when $\kappa_d > K_0$ and $\kappa_d \le K_0$ respectively. For the simplicity of notation, define two indicator functions $I_1 = \mathbf{1}_{\{|j_1-j_2|\le K_0, |j_3-j_4|\le K_0, \kappa_d > K_0\}}$ and $I_2 = \mathbf{1}_{\{|j_1-j_2|\le K_0, |j_3-j_4|\le K_0, \kappa_d \le K_0\}}$. By definition, we have $\mathbf{X}_{j_1,j_2,j_3,j_4} = \mathrm{cov}(x_{1,j_1}x_{1,j_2}, x_{1,j_3}x_{1,j_4})$ and $\mathbf{Y}_{j_1,j_2,j_3,j_4} = \mathrm{cov}(y_{1,j_1}y_{1,j_2}, y_{1,j_3}y_{1,j_4})$. When $\kappa_d > K_0$, we know $\mathbf{X}_{j_1,j_2,j_3,j_4} \le C\delta^{\frac{K_0\epsilon}{2+\epsilon}}$ by Condition 3.4.1 (2) and (3) and Lemma B.5.1. It follows that

$$\begin{aligned} &\Big| \sum_{1\le j_1,j_2,j_3,j_4\le p} (\mathbf{X}_{j_1,j_2,j_3,j_4})^m (\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m} \times I_1 \Big| \\ \le & C p^4 \delta^{\frac{K_0\epsilon}{2+\epsilon}} = O(1) p^4 \times p^{-(8+2\mu)} = o(1). \end{aligned} \tag{B.183}$$

In addition, note that $\sum_{1\le j_1,j_2,j_3,j_4\le p} I_2 = O(pK_0^3) = O(p\log^3 p)$. By Condition 3.4.1 (2), we know

$$\Big| \sum_{1\le j_1,j_2,j_3,j_4\le p} (\mathbf{X}_{j_1,j_2,j_3,j_4})^m (\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m} \times I_2 \Big| = O(p\log^3 p).$$

Case 2: If $|j_1 - j_2| > K_0$ or $|j_3 - j_4| > K_0$, by Lemma B.5.1, we know that $|\sigma_{j_1,j_2}\sigma_{j_3,j_4}| \le C\delta^{\frac{K_0\epsilon}{2+\epsilon}}$. We consider $|j_1 - j_2| > K_0$ without loss of generality and discuss the following cases (i)–(iv).

(i) When $|j_2 - j_3| > K_0/2$ and $|j_2 - j_4| > K_0/2$,

$$|\mathbf{X}_{j_1,j_2,j_3,j_4}| = |\mathrm{cov}(x_{1,j_1}x_{1,j_3}x_{1,j_4}\,,\, x_{1,j_2}) - \sigma_{j_1,j_2}\sigma_{j_3,j_4}| \le C\delta^{\frac{K_0\epsilon}{2(2+\epsilon)}}.$$

(ii) When $|j_2 - j_3| \le K_0/2$ and $|j_2 - j_4| \le K_0/2$, we know that $|j_1 - j_3| \ge$

$|j_1 - j_2| - |j_2 - j_3| > K_0/2$ and $|j_1 - j_4| \geq |j_1 - j_2| - |j_2 - j_4| > K_0/2$. Then

$$|\mathbf{X}_{j_1,j_2,j_3,j_4}| = |\mathrm{cov}(x_{1,j_1}, x_{1,j_2}x_{1,j_3}x_{1,j_4}) - \sigma_{j_1,j_2}\sigma_{j_3,j_4}| \leq C\delta^{\frac{K_0\epsilon}{2(2+\epsilon)}}. \tag{B.184}$$

(iii) When $|j_2 - j_3| \leq K_0/2$ and $|j_2 - j_4| > K_0/2$, as we know $|j_1 - j_2| > K_0$, then $|j_1 - j_3| > K_0/2$. We next discuss three sub-cases.

(iiia) If $|j_1 - j_4| > K_0/2$, we know (B.184) also holds.

For easy presentation, let I_3 be an indicator function when $\{j_1, j_2, j_3, j_4\}$ satisfies the sub-cases (i), (ii) and (iiia) above. Then similarly to (B.183),

$$\Big| \sum_{1 \leq j_1,j_2,j_3,j_4 \leq p} (\mathbf{X}_{j_1,j_2,j_3,j_4})^m (\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m} \times I_3 \Big| = o(1).$$

(iiib) If $|j_1 - j_4| \leq K_0/2$, and $|j_3 - j_4| \leq K_0/2$, we know under this case $|j_2 - j_3|, |j_1 - j_4|, |j_3 - j_4| \leq K_0$. Let $I_4 = \mathbf{1}_{\{|j_2-j_3|,|j_1-j_4|,|j_3-j_4| \leq K_0\}}$. We have $\sum_{1 \leq j_1,j_2,j_3,j_4 \leq p} I_4 = O(pK_0^3)$. By Condition 3.4.1 (2), we know

$$\Big| \sum_{1 \leq j_1,j_2,j_3,j_4 \leq p} (\mathbf{X}_{j_1,j_2,j_3,j_4})^m (\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m} \times I_4 \Big| = O(p \log^3 p).$$

(iiic) If $|j_1 - j_4| \leq K_0/2$, and $|j_3 - j_4| > K_0/2$, we know

$$\mathbf{X}_{j_1,j_2,j_3,j_4} \geq \mathrm{E}(x_{1,j_1}x_{1,j_4})\mathrm{E}(x_{1,j_2}x_{1,j_3}) - C\delta^{\frac{K_0\epsilon}{2(2+\epsilon)}}.$$

Let I_5 be an indicator function of the sub-case (iiic) above. Then

$$\begin{aligned}
&\Big|\sum_{1\le j_1,j_2,j_3,j_4\le p}(\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m}\times I_5\Big| \\
&=\Big|\sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_4}\sigma_{j_2,j_3})^a\times I_5\Big|+O(p^4p^{-(4+\mu)}) \\
&=\Big|\sum_{|j_1-j_4|\le K_0/2,\,|j_2-j_3|\le K_0/2}(\sigma_{j_1,j_4}\sigma_{j_2,j_3})^a\Big|+o(1) \\
&=\Big|\sum_{1\le j_1,j_2,j_3,j_4\le p}(\sigma_{j_1,j_4}\sigma_{j_2,j_3})^a-\sum_{|j_1-j_4|>K_0\text{ or }|j_2-j_3|>K_0}(\sigma_{j_1,j_4}\sigma_{j_2,j_3})^a\Big|+o(1) \\
&=\Theta(p^2).
\end{aligned}$$

where the last equation uses Conditions 3.4.1 (3) and (4) and Lemma B.5.1.

(iv) When $|j_2-j_3|>K_0/2$ and $|j_2-j_4|\le K_0/2$, this is symmetric to the sub-case (iii) discussed above. Define an indicator function $I_6=\mathbf{1}_{\{|j_2-j_3|>K_0/2,|j_2-j_4|\le K_0/2\}}$. We then have

$$\Big|\sum_{1\le j_1,j_2,j_3,j_4\le p}(\mathbf{X}_{j_1,j_2,j_3,j_4})^m(\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m}\times I_6\Big|=\Theta(p^2).$$

In summary, (B.182) is proved and thus $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}=\Theta(p^2n^{-a})$ is obtained. To prove $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}=o(1)\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$, it remains to show that $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}=o(p^2n^{-a})$.

We write $\mathcal{U}(a)=\sum_{c=0}^{a}\sum_{b_1=0}^{c}\sum_{b_2=0}^{a-c}C_{a,c,b_1,b_2}T_{b_1,b_2,c}$, where we define $C_{a,c,b_1,b_2}=(-1)^{c-b_1+b_2}a!/\{b_1!b_2!(c-b_1)!(a-c-b_2)!\}$, and

$$\begin{aligned}
T_{b_1,b_2,c}=&\sum_{1\le j_1,j_2\le p}\sum_{\substack{\mathbf{i}\in\mathcal{P}(n_x,2c-b_1);\\ \mathbf{w}\in\mathcal{P}(n_y,2(a-c)-b_2)}}(P^{n_x}_{2c-b_1}P^{n_y}_{2(a-c)-b_2})^{-1} \\
&\times\prod_{k=1}^{b_1}(x_{i_k,j_1}x_{i_k,j_2}-\sigma_{j_1,j_2})\prod_{k=b_1+1}^{c}x_{i_k,j_1}\prod_{k=c+1}^{2c-b_1}x_{i_k,j_2} \\
&\times\prod_{m=1}^{b_2}(y_{w_m,j_1}y_{w_m,j_2}-\sigma_{j_1,j_2})\prod_{l=b_2+1}^{a-c}y_{w_l,j_1}\prod_{q=a-c+1}^{2(a-c)-b_2}y_{w_q,j_2}.
\end{aligned}\tag{B.185}$$

Then $\tilde{\mathcal{U}}(a) = \sum_{c=0}^{a}(-1)^{a-c}T_{c,a-c,c}$ and $\tilde{\mathcal{U}}^*(a) = \sum_{c=0}^{a}\sum_{b_1=0}^{c}\sum_{b_2=0}^{a-c} C_{a,c,b_1,b_2}T_{b_1,b_2,c} \times \mathbf{1}_{b_1+b_2\leq a-1}$. Note that $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} \leq C\max_{b_1,b_2,c;b_1+b_2\leq a-1}\{\mathrm{var}(T_{b_1,b_2,c})\}$, where C is some constant. When a is finite, to prove $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(p^2n^{-a})$, it suffices to show that $\mathrm{var}(T_{b_1,b_2,c}) = o(p^2n^{-a})$ for each (b_1, b_2, c) satisfying $b_1 + b_2 \leq a - 1$. Note that $\mathrm{E}(T_{b_1,b_2,c}) = 0$ under H_0, then $\mathrm{var}(T_{b_1,b_2,c}) = \mathrm{E}(T^2_{b_1,b_2,c})$ and

$$\mathrm{var}(T_{b_1,b_2,c}) = (P^{n_x}_{2c-b_1}P^{n_y}_{2(a-c)-b_2})^{-2}\sum_{\substack{1\leq j_1,j_2\leq p;\\ 1\leq\tilde{j}_1,\tilde{j}_2\leq p}}\ \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,2c-b_1);\\ \mathbf{w}\,\tilde{\mathbf{w}}\in\mathcal{P}(n_y,2(a-c)-b_2)}} \mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2), \tag{B.186}$$

where we let

$$\begin{aligned}
&\mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2)\\
&= \mathrm{E}\Big\{\prod_{k=1}^{b_1}(x_{i_k,j_1}x_{i_k,j_2}-\sigma_{j_1,j_2})(x_{\tilde{i}_k,\tilde{j}_1}x_{\tilde{i}_k,\tilde{j}_2}-\sigma_{\tilde{j}_1,\tilde{j}_2})\prod_{k=b_1+1}^{c}(x_{i_k,j_1}x_{\tilde{i}_k,\tilde{j}_1})\\
&\quad\times\prod_{k=c+1}^{2c-b_1}(x_{i_k,j_2}x_{\tilde{i}_k,\tilde{j}_2})\Big\}\mathrm{E}\Big\{\prod_{m=1}^{b_2}(y_{w_m,j_1}y_{w_m,j_2}-\sigma_{j_1,j_2})(y_{\tilde{w}_m,\tilde{j}_1}y_{\tilde{w}_m,\tilde{j}_2}-\sigma_{j_1,j_2})\\
&\quad\times\prod_{m=b_2+1}^{a-c}(y_{w_m,j_1}y_{\tilde{w}_m,\tilde{j}_1})\prod_{m=a-c+1}^{2(a-c)-b_2}(y_{w_m,j_2}y_{\tilde{w}_m,\tilde{j}_2})\Big\}.
\end{aligned}$$

Since we assume without loss of generality that $\mathrm{E}(\mathbf{x}) = \mathrm{E}(\mathbf{y}) = \mathbf{0}$, then $\mathrm{E}(x_{1,j_1}x_{1,j_2} - \sigma_{j_1,j_2}) = \mathrm{E}(y_{1,j_1}x_{1,j_2} - \sigma_{j_1,j_2}) = 0$. It follows that when $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$ or $\{\mathbf{w}\} \neq \{\tilde{\mathbf{w}}\}$, $\mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2) = 0$. When $\{\mathbf{i}\} = \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{w}\} = \{\tilde{\mathbf{w}}\}$, we have $|\{\mathbf{i}\}\cup\{\tilde{\mathbf{i}}\}| + |\{\mathbf{w}\}\cup\{\tilde{\mathbf{w}}\}| = 2c - b_1 + 2(a-c) - b_2$. By Condition 3.4.1 (1) and (2), for any given $\{j_1, j_2, \tilde{j}_1, \tilde{j}_2\}$,

$$\begin{aligned}
&(P^{n_x}_{2c-b_1}P^{n_y}_{2(a-c)-b_2})^{-2}\sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,2c-b_1);\\ \mathbf{w}\,\tilde{\mathbf{w}}\in\mathcal{P}(n_y,2(a-c)-b_2)}}\mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2) \qquad \text{(B.187)}\\
&= O(n^{-2(2a+b_1+b_2)}\times n^{2a-b_1-b_2}) = O(n^{-2a+b_1+b_2}) = o(n^{-a-1})
\end{aligned}$$

where in the last equation, we use $b_1 + b_2 \le a - 1$. In addition, similarly to (B.182), we have that for any given $(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}})$,

$$\sum_{1 \le j_1, j_2, \tilde{j}_1, \tilde{j}_2 \le p} \mathbb{T}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, j_1, j_2, \tilde{j}_1, \tilde{j}_2) = O(p^2). \tag{B.188}$$

In summary, by (B.187) and (B.188), we know $\text{var}\{\tilde{\mathcal{U}}^*(a)\} = O(p^2 n^{-a-1}) = o(p^2 n^{-a})$.

B.5.23.2 Proof under Condition 3.4.1*

In this section, we prove that $\text{var}\{\tilde{\mathcal{U}}(a)\} = o(1)\text{var}\{\tilde{\mathcal{U}}^*(a)\}$ under Condition 3.4.1*. Recall that we have already obtained $\text{var}\{\tilde{\mathcal{U}}(a)\}$ in (B.181). By Condition 3.4.1* (3), we have

$$\begin{aligned} \mathbf{X}_{j_1,j_2,j_3,j_4} &= \kappa_x(\sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3}) + (\kappa_x - 1)\sigma_{j_1,j_2}\sigma_{j_3,j_4}, \\ \mathbf{Y}_{j_1,j_2,j_3,j_4} &= \kappa_y(\sigma_{j_1,j_3}\sigma_{j_2,j_4} + \sigma_{j_1,j_4}\sigma_{j_2,j_3}) + (\kappa_y - 1)\sigma_{j_1,j_2}\sigma_{j_3,j_4}. \end{aligned} \tag{B.189}$$

Then by Condition 3.4.1* (1) and (4), we know $(\mathbf{X}_{j_1,j_2,j_3,j_4})^m (\mathbf{Y}_{j_1,j_2,j_3,j_4})^{a-m}$ is a linear combination of

$$\prod_{t=1}^{a} \left\{ \sigma_{j_{g_1^{(t)}}, j_{g_2^{(t)}}} \times \sigma_{j_{g_3^{(t)}}, j_{g_4^{(t)}}} \right\}, \tag{B.190}$$

where $\{(g_1^{(t)}, g_2^{(t)}), (g_3^{(t)}, g_4^{(t)}) : t = 1, \ldots, a\}$ are a allocations of the set $\{1, 2, 3, 4\}$ into 2 (unordered) pairs. When the a allocations are the same, by the symmetricity of j indexes,

$$\sum_{1 \le j_1, j_2, j_3, j_4 \le p} \prod_{t=1}^{a} \sigma_{j_{g_1^{(t)}}, j_{g_2^{(t)}}} \sigma_{j_{g_3^{(t)}}, j_{g_4^{(t)}}} = \sum_{1 \le j_1, j_2, j_3, j_4 \le p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a.$$

When the a allocations are different, by Condition 3.4.1* (4),

$$\sum_{1\leq j_1,j_2,j_3,j_4\leq p} \prod_{t=1}^{a} \sigma_{j_{g_1^{(t)}}, j_{g_2^{(t)}}} \sigma_{j_{g_3^{(t)}}, j_{g_4^{(t)}}} = o(1) \sum_{1\leq j_1,j_2,j_3,j_4\leq p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a, \tag{B.191}$$

which can be obtained by taking square of both sides of (B.191) and using Condition 3.4.1* (4). It follows that by (B.181), Condition 3.4.1* (1) and (4) and the symmetricity of j indexes,

$$\mathrm{var}\{\tilde{\mathcal{U}}(a)\} = \Theta(n^{-a}) \sum_{1\leq j_1,j_2,j_3,j_4\leq p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a. \tag{B.192}$$

We next show $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$. Similarly to Section B.5.23.1, we know it suffices to prove $\mathrm{var}(T_{b_1,b_2,c}) = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ for $0 \leq c \leq a$, $0 \leq b_1 \leq c$, $0 \leq b_2 \leq a-c$ and $b_1 + b_2 \leq a-1$. Note that (B.186) still holds here, and when $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$ or $\{\mathbf{w}\} \neq \{\tilde{\mathbf{w}}\}$, $\mathbb{T}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, j_1, j_2, \tilde{j}_1, \tilde{j}_2) = 0$. Therefore, (B.187) also holds. By Condition 3.4.1* (3) and (4), similarly to the analysis of (B.192), we have for any given $(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}})$,

$$\sum_{1\leq j_1,j_2,j_3,j_4\leq p} \mathbb{T}(\mathbf{i}, \tilde{\mathbf{i}}, \mathbf{w}, \tilde{\mathbf{w}}, j_1, j_2, \tilde{j}_1, \tilde{j}_2) = O(1) \sum_{1\leq j_1,j_2,j_3,j_4\leq p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a \tag{B.193}$$

Combining (B.187) and (B.193),

$$\mathrm{var}(T_{b_1,b_2,c}) = O(n^{-a-1}) \sum_{1\leq j_1,j_2,j_3,j_4\leq p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a = o(1)\mathrm{var}\{\tilde{\mathcal{U}}(a)\}.$$

B.5.24 Proof of Lemma B.3.2

As $\mathrm{E}\{\mathcal{U}(a)\} = \mathrm{E}\{\mathcal{U}(b)\} = 0$ under H_0, we know that $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} = \mathrm{E}\{\mathcal{U}(a)\mathcal{U}(b)\}/\{\sigma(a)\sigma(b)\}$. Recall that $\mathcal{U}(a) = \tilde{\mathcal{U}}(a) + \tilde{\mathcal{U}}^*(a)$ and $\mathcal{U}(b) = \tilde{\mathcal{U}}(b) + \tilde{\mathcal{U}}^*(b)$.

Then

$$\mathrm{E}\Big\{\frac{\mathcal{U}(a)}{\sigma(a)}\times\frac{\mathcal{U}(b)}{\sigma(b)}\Big\} = \mathrm{E}\Big\{\frac{\tilde{\mathcal{U}}(a)+\tilde{\mathcal{U}}^*(a)}{\sigma(a)}\times\frac{\tilde{\mathcal{U}}(b)+\tilde{\mathcal{U}}^*(b)}{\sigma(b)}\Big\} \tag{B.194}$$
$$= \mathrm{E}\Big\{\frac{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)}{\sigma(a)\sigma(b)}\Big\} + o(1),$$

where the last equation follows by Lemma B.3.1. By the definition and notation in Section B.5.23,

$$\tilde{\mathcal{U}}(a) = \tilde{C}_a \sum_{\substack{1\le j_1,j_2\le p;\\ \mathbf{i}\in\mathcal{P}(n_x,a);\\ \mathbf{w}\in\mathcal{P}(n_y,a)}} \mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2), \quad \tilde{\mathcal{U}}(b) = \tilde{C}_b \sum_{\substack{1\le \tilde{j}_1,\tilde{j}_2\le p;\\ \tilde{\mathbf{i}}\in\mathcal{P}(n_x,b);\\ \tilde{\mathbf{w}}\in\mathcal{P}(n_y,b)}} \mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},\tilde{j}_1,\tilde{j}_2),$$

where we let $\tilde{C}_a = (P^{n_x}_a P^{n_y}_a)^{-1}$, $\tilde{C}_b = (P^{n_x}_b P^{n_y}_b)^{-1}$, $\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2) = \prod_{t=1}^a (x_{i_t,j_1}x_{i_t,j_2} - y_{w_t,j_1}y_{w_t,j_2})$ and $\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},\tilde{j}_1,\tilde{j}_2) = \prod_{t=1}^b (x_{\tilde{i}_t,\tilde{j}_1}x_{\tilde{i}_t,\tilde{j}_2} - y_{\tilde{w}_t,\tilde{j}_1}y_{\tilde{w}_t,\tilde{j}_2})$. It follows that

$$\mathrm{E}\{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)\} = \tilde{C}_a\tilde{C}_b \sum_{\substack{1\le j_1,j_2,\tilde{j}_1,\tilde{j}_2\le p;\\ \mathbf{i}\in\mathcal{P}(n_x,a);\,\tilde{\mathbf{i}}\in\mathcal{P}(n_x,b)\\ \mathbf{w}\in\mathcal{P}(n_y,a);\,\tilde{\mathbf{w}}\in\mathcal{P}(n_y,b)}} \mathrm{E}\Big\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},\tilde{j}_1,\tilde{j}_2)\Big\}.$$

As $a \neq b$, we know $\{\mathbf{i}\} \neq \{\tilde{\mathbf{i}}\}$ and $\{\mathbf{w}\} \neq \{\tilde{\mathbf{w}}\}$. It follows that similarly to Section B.5.3, $\mathrm{E}\{\mathbb{D}_{\mathbf{x},\mathbf{y}}(\mathbf{i},\mathbf{w},j_1,j_2)\mathbb{D}_{\mathbf{x},\mathbf{y}}(\tilde{\mathbf{i}},\tilde{\mathbf{w}},\tilde{j}_1,\tilde{j}_2)\} = 0$. Therefore $\mathrm{E}\{\tilde{\mathcal{U}}(a)\tilde{\mathcal{U}}(b)\} = 0$ and $\mathrm{cov}\{\mathcal{U}(a)/\sigma(a), \mathcal{U}(b)/\sigma(b)\} = o(1)$.

B.5.25 Deriving $D_{n,k}$ and $\pi^2_{n,k}$ in Lemmas B.3.3 and B.3.4

To prove Lemmas B.3.3 and B.3.4, we derive the forms of $D_{n,k}$ and $\pi^2_{n,k}$ in this section. By construction, $D_{n,k} = \sum_{r=1}^m t_r A_{n,k,a_r}$, where $A_{n,k,a_r} = (\mathrm{E}_k - \mathrm{E}_{k-1})[\tilde{\mathcal{U}}(a_r)/\sigma(a_r)]$. In addition, $\pi^2_{n,k} = \sum_{1\le r_1,r_2\le m} t_{r_1}t_{r_2}\mathrm{E}_{k-1}(A_{n,k,a_{r_1}}A_{n,k,a_{r_2}})$. It then suffices to derive the form of $A_{n,k,a}$ for a given integer a, and also derive $\mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})$ for two given integers a_1 and a_2.

For easy presentation, we define $\mathcal{X}_{i,j_1,j_2} = x_{i,j_1}x_{i_t,j_2} - \sigma_{j_1,j_2}$ and $\mathcal{Y}_{i,j_1,j_2} = y_{i,j_1}y_{i_t,j_2} -$

σ_{j_1,j_2} in the following. Then under H_0,

$$\tilde{\mathcal{U}}(a) = (P_a^{n_x} P_a^{n_y})^{-1} \sum_{\substack{1\le j_1,j_2\le p;\\ \mathbf{i}\in\mathcal{P}(n_x,a);\, \mathbf{w}\in\mathcal{P}(n_y,a)}} \prod_{t=1}^{a} (\mathcal{X}_{w_t,j_1,j_2} - \mathcal{Y}_{i_t,j_1,j_2}).$$

Part I: $1 \le k \le n_x$ When $1 \le k \le n_x$, similarly to Section B.5.5, as $\mathrm{E}(\mathcal{X}_{1,j_1,j_2}) = 0$ under H_0, we have

$$(\mathrm{E}_k - \mathrm{E}_{k-1})\Big\{ \prod_{t=1}^{a} (\mathcal{X}_{i_t,j_1,j_2} - \mathcal{Y}_{w_t,j_1,j_2}) \Big\} = (\mathrm{E}_k - \mathrm{E}_{k-1})\Big(\prod_{t=1}^{a} \mathcal{X}_{i_t,j_1,j_2} \Big),$$

which is nonzero only when $i_1, \ldots, i_a \le k$ and $k \in \{i_1, \ldots, i_a\}$. Then we know when $k < a$, $A_{n,k,a} = 0$ and when $k \ge a$,

$$A_{n,k,a} = c_1(n,a) \sum_{\substack{1\le j_1,j_2\le p;\\ \mathbf{i}\in\mathcal{P}(k-1,a-1)}} \Big(\prod_{t=1}^{a-1} \mathcal{X}_{i_t,j_1,j_2} \Big) \mathcal{X}_{k,j_1,j_2}, \tag{B.195}$$

where $c_1(n,a) = a!/\{P_a^{n_x}\sigma(a)\}$. For two integers a_1 and a_2,

$$\begin{aligned} &\mathrm{E}_{k-1}(A_{n,k,a_1} A_{n,k,a_2}) \\ &= \prod_{l=1}^{2} c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2}} \mathbb{M}_{\mathbf{x},\mathbf{y},1}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2), \end{aligned}$$

where

$$\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1,2) = \prod_{l=1}^{2} \Big(\prod_{t=1}^{a_l-1} \mathcal{X}_{i_t^{(l)},j_{2l-1},j_{2l}} \Big) \mathrm{E}(\mathcal{X}_{k,j_1,j_2} \mathcal{X}_{k,j_3,j_4}).$$

Part II: $n_x + 1 \le k \le n_x + n_y$ When $n_x + 1 \le k \le n_x + n_y$, we have

$$\prod_{t=1}^{a} (\mathcal{X}_{i_t,j_1,j_2} - \mathcal{Y}_{i_t,j_1,j_2}) = \sum_{s=0}^{a} \sum_{\substack{\mathbf{i}^*\in\mathcal{S}(\mathbf{i},s);\\ \mathbf{w}^*\in\mathcal{S}(\mathbf{w},a-s)}} \Big(\prod_{t=1}^{s} \mathcal{X}_{i_t^*,j_1,j_2} \Big) \Big(\prod_{\tilde{t}=1}^{a-s} \mathcal{Y}_{w_{\tilde{t}}^*,j_1,j_2} \Big),$$

where $\mathcal{S}(\mathbf{i}, s)$ represents the collection of sub-tuples of $\mathbf{i}$ with length s and $\mathcal{S}(\mathbf{w}, a-s)$ represents the collection of sub-tuples of $\mathbf{w}$ with length $a-s$, which is similarly used in Section B.5.11. When $n_x+1 \leq k \leq n_x+n_y$, similarly to Section B.5.5, $(\mathrm{E}_k - \mathrm{E}_{k-1})\{\prod_{t=1}^{s}(x_{i_t^*,j_1}x_{i_t^*,j_2} - \sigma_{j_1,j_2})\prod_{\tilde{t}=1}^{a-s}(y_{w_{\tilde{t}}^*,j_1}y_{w_{\tilde{t}}^*,j_2} - \sigma_{j_1,j_2})\} \neq 0$ only when $w_1^*, \ldots, w_{a-s}^* \leq k-n_x$ and $k-n_x \in \{w_1^*, \ldots, w_{a-s}^*\}$, and then

$$(\mathrm{E}_k - \mathrm{E}_{k-1})\Big(\prod_{t=1}^{s}\mathcal{X}_{i_t^*,j_1,j_2}\prod_{\tilde{t}=1}^{a-s}\mathcal{Y}_{w_{\tilde{t}}^*,j_1,j_2}\Big) = \mathcal{Y}_{k-n_x,j_1,j_2}\prod_{t=1}^{s}\mathcal{X}_{i_t^*,j_1,j_2}\prod_{\tilde{t}=1}^{a-s-1}\mathcal{Y}_{w_{\tilde{t}}^*,j_1,j_2}.$$

It follows that

$$A_{n,k,a} = \sum_{s=L_k}^{a-1}\sum_{\substack{1\leq j_1,j_2\leq p;\\ \mathbf{i}\in\mathcal{P}(n_x,s);\\ \mathbf{w}\in\mathcal{P}(k-n_x-1,a-s-1)}} c_2(n,a,s)\mathcal{Y}_{k-n_x,j_1,j_2}\prod_{t=1}^{s}\mathcal{X}_{i_t,j_1,j_2}\prod_{\tilde{t}=1}^{a-s-1}\mathcal{Y}_{w_{\tilde{t}},j_1,j_2},$$

where $L_k = \max\{n_x - k + a, 0\}$ and $c_2(n,a,s) = P_{a-s}^{n_x-s}P_s^{n_y-a+s}\{P_a^{n_x}P_a^{n_y}\sigma(a)\}^{-1}$. Thus for two constants a_1 and a_2,

$$\begin{aligned}&\mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})\\ =&\sum_{\substack{1\leq j_1,j_2,j_3,j_4\leq p;\\ L_k\leq s_l\leq a_l:\,l=1,2;\\ \mathbf{i}^{(l)}\in\mathcal{P}(n_x,s_l):\,l=1,2;\\ \mathbf{w}^{(l)}\in\mathcal{P}(k-n_x-1,a_l-s_l-1):\,l=1,2}}\prod_{l=1}^{2}c_2(n,a_l,s_l)\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2),\end{aligned}$$

where

$$\begin{aligned}&\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\\ =&\prod_{l=1}^{2}\Big(\prod_{t=1}^{s_l}\mathcal{X}_{i_t^{(l)},j_{2l-1},j_{2l}}\prod_{\tilde{t}=1}^{a_l-s_l-1}\mathcal{Y}_{w_{\tilde{t}}^{(l)},j_{2l-1},j_{2l}}\Big)\mathrm{E}(\mathcal{Y}_{k-n_x,j_1,j_2}\mathcal{Y}_{k-n_x,j_3,j_4}).\end{aligned}$$

B.5.26 Proof of Lemma B.3.3

Note that by the Cauchy-Schwarz inequality, for some constant C,

$$\operatorname{var}\Big(\sum_{k=1}^{n}\pi_{n,k}^2\Big) \le Cn^2 \max_{1\le k\le n;\, 1\le r_1,r_2\le m} \operatorname{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}}),$$

where for two integers a_1 and a_2, $\mathbb{T}_{k,a_1,a_2} = \mathrm{E}_{k-1}(A_{n,k,a_1}A_{n,k,a_2})$ is given in Section B.5.25. Therefore to prove Lemma B.3.3, it suffices to prove $\operatorname{var}(\mathbb{T}_{k,a_{r_1},a_{r_2}}) = o(n^{-2})$ for every $1 \le k \le n$ and $1 \le r_1, r_2 \le m$. We next prove $\operatorname{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$ when $a \le k \le n_x$ and $n_x + 1 \le k \le n_x + n_y$ in the following Parts I and II respectively.

Part I: $a \le k \le n_x$ We first derive the form of $\operatorname{var}(\mathbb{T}_{k,a_1,a_2})$ when $a \le k \le n_x$. As $\operatorname{var}(\mathbb{T}_{k,a_1,a_2}) = \mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2$, we next derive $\mathrm{E}(\mathbb{T}_{k,a_1,a_2})$ and $\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2)$. In particular,

$$\mathrm{E}(\mathbb{T}_{k,a_1,a_2}) = \prod_{l=1}^{2} c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2}} \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2)\Big\}.$$

For easy presentation, we let $a_3 = a_1$ and $a_4 = a_2$, and have

$$\begin{aligned}
&\Big\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\Big\}^2 \\
=\;& \prod_{l=1}^{4} c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2)\Big\} \\
&\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=3,4)\Big\}.
\end{aligned}$$

In addition, we have

$$\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) = \prod_{l=1}^{4} c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\},$$

where we define

$$\begin{aligned}&\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\\ =&\prod_{l=1}^{4}\Big(\prod_{t=1}^{a_l-1}\mathcal{X}_{i_t^{(l)},j_{2l-1},j_{2l}}\Big)\mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_3,j_4})\mathrm{E}(\mathcal{X}_{k,j_5,j_6}\mathcal{X}_{k,j_7,j_8}).\end{aligned}$$

Let $\mathbf{1}_E$ be an indicator function of the event that $(\{\mathbf{i}^{(1)}\}\cup\{\mathbf{i}^{(2)}\})\cap(\{\mathbf{i}^{(3)}\}\cup\{\mathbf{i}^{(4)}\}) = \emptyset$. Then define

$$G_{a_1,a_2,1} = \prod_{l=1}^{4} c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} \times\mathbf{1}_E \times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\}.$$

We also note that

$$\begin{aligned}&\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\}\times\mathbf{1}_E \\ =\ &\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2)\Big\}\\ &\times\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=3,4)\Big\}\times\mathbf{1}_E.\end{aligned} \tag{B.196}$$

Since $|\mathrm{var}(\mathbb{T}_{k,a_1,a_2})| \le |\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - G_{a_1,a_2,1}| + |\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}|$, to prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$, we will next show that $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| = o(n^{-2})$ and $|\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - G_{a_1,a_2,1}| = o(n^{-2})$. In particular, we present the proof under Conditions 3.4.1 and 3.4.1* in the following Sections B.5.26 and B.5.26, respectively.

Proof under Condition 3.4.1

Step I: $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| = o(n^{-2})$. If $a_1 \neq a_2$, we have $\mathrm{E}(\mathbb{T}_{k,a_1,a_2}) = G_{a_1,a_2,1} = 0$. It remains to consider $a_1 = a_2$ below. Note that

$$\begin{aligned}
&\mathrm{E}\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2)\} \\
&\times \mathrm{E}\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 3, 4)\}
\end{aligned} \tag{B.197}$$

satisfies that (B.197) $\neq 0$ only if $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$ and $\{\mathbf{i}^{(3)}\} = \{\mathbf{i}^{(4)}\}$. Thus,

$$\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 = \prod_{l=1}^{4} c(n, a_l) \sum_{\substack{1 \le j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8 \le p; \\ \mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1),\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}, \\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}}\right\}} \times \text{(B.197)}.$$

Similarly, $\mathrm{E}\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\} \times \mathbf{1}_E \neq 0$ only when $\{\mathbf{i}^{(1)}\} = \{\mathbf{i}^{(2)}\}$ and $\{\mathbf{i}^{(3)}\} = \{\mathbf{i}^{(4)}\}$. Therefore, by (B.196),

$$G_{a_1,a_2,1} = \prod_{l=1}^{4} c(n, a_l) \sum_{\substack{1 \le j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8 \le p; \\ \mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1),\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}, \\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}, \\ \{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}=\emptyset}\right\}} \times \text{(B.197)},$$

and then

$$\begin{aligned}
&|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| \\
\le\ &\prod_{l=1}^{4} c(n, a_l) \sum_{\substack{1 \le j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8 \le p; \\ \mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1),\, l=1,2,3,4}} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}, \\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}, \\ \{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}\neq\emptyset}\right\}} \times |\text{(B.197)}|.
\end{aligned} \tag{B.198}$$

Note that

$$\sum_{\mathbf{i}^{(l)} \in \mathcal{P}(k-1, a_l - 1),\, l=1,2,3,4} \mathbf{1}_{\left\{\substack{\{\mathbf{i}^{(1)}\}=\{\mathbf{i}^{(2)}\}, \\ \{\mathbf{i}^{(3)}\}=\{\mathbf{i}^{(4)}\}, \\ \{\mathbf{i}^{(1)}\}\cap\{\mathbf{i}^{(3)}\}\neq\emptyset}\right\}} = O(n^{a_1+a_2-3}). \tag{B.199}$$

In addition, by Condition 3.4.1 (2),

$$
\begin{aligned}
&\sum_{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p} |(\text{B.197})| \\
\le& C \sum_{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p} \Big| \mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_3,j_4})\mathrm{E}(\mathcal{X}_{k,j_5,j_6}\mathcal{X}_{k,j_7,j_8}) \Big|.
\end{aligned} \tag{B.200}
$$

Recall that $\mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_3,j_4}) = \mathbf{X}_{j_1,j_2,j_3,j_4}$ and $\mathrm{E}(\mathcal{X}_{k,j_5,j_6}\mathcal{X}_{k,j_7,j_8}) = \mathbf{X}_{j_5,j_6,j_7,j_8}$ following the notation in Section B.5.23. Following the similar analysis for the proof of (B.182), we obtain $\sum_{1\le j_1,j_2,j_3,j_4\le p} |\mathbf{X}_{j_1,j_2,j_3,j_4}| = O(p^2)$ and $\sum_{1\le j_5,j_6,j_7,j_8\le p} |\mathbf{X}_{j_5,j_6,j_7,j_8}| = O(p^2)$. It follows that (B.200) $= O(p^4)$. Note that $c(n,a) = \Theta(p^{-1}n^{-a/2})$ by Lemma B.3.1. Combining (B.199) and (B.200), we obtain $\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1} = o(n^{-2})$.

Step II: $|\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{a_1,a_2,1}| = o(n^{-2})$. By construction, we have

$$
\begin{aligned}
\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{a_1,a_2,1} = \prod_{l=1}^{4} c(n,a_l) &\sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} (1-\mathbf{1}_E) \\
&\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2,3,4)\Big\}.
\end{aligned} \tag{B.201}
$$

When $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| > a_1+a_2-2$, which means that there exists one index that only appears once among the four sets $\{\mathbf{i}^{(l)}\}$, $l=1,2,3,4$, then similarly to Section B.5.6,

$$
\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}: l=1,2)\Big\} \times (1-\mathbf{1}_E) \tag{B.202}
$$

satisfies that (B.202) $= 0$. When $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| < a_1+a_2-2$,

$$
\sum_{\mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4} \mathbf{1}_{\{|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|<a_1+a_2-2\}} = O(n^{a_1+a_2-3}). \tag{B.203}
$$

Similarly to the analysis of (B.200) above, by Condition 3.4.1, we have

$$\sum_{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p} (\text{B.202}) = O(p^4). \tag{B.204}$$

Therefore, by (B.203), (B.204) and $c(n,a) = \Theta(p^{-1}n^{-a/2})$,

$$\begin{aligned} &\prod_{l=1}^4 c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} (1-\mathbf{1}_E)\mathbf{1}_{\{|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|<a_1+a_2-2\}} \\ &\qquad\qquad \times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\Big\} \\ =\ & O(1)n^{-a_1-a_2}p^{-4}n^{a_1+a_2-3}p^4 = o(n^{-2}). \end{aligned}$$

Last, we consider $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| = a_1+a_2-2$. Note that $1-\mathbf{1}_E\neq 0$ indicates that $(\{\mathbf{i}^{(1)}\}\cup\{\mathbf{i}^{(2)}\})\cap(\{\mathbf{i}^{(3)}\}\cup\{\mathbf{i}^{(4)}\})\neq\emptyset$ under this case. By the symmetricity of the j indexes, we have

$$\begin{aligned} &\Big|\sum_{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p}(\text{B.202})\Big| \qquad\qquad (\text{B.205})\\ \le\ & C\sum_{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p}\Big|\mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_3,j_4})\mathrm{E}(\mathcal{X}_{k,j_5,j_6}\mathcal{X}_{k,j_7,j_8}) \\ &\qquad \times \mathrm{E}(\mathcal{X}_{k,j_1,j_2}\mathcal{X}_{k,j_5,j_6})\mathrm{E}(\mathcal{X}_{k,j_3,j_4}\mathcal{X}_{k,j_7,j_8})\Big|. \end{aligned}$$

Following similar arguments to that in Sections B.5.6.1 and B.5.23.1, by discussing different cases of j indexes, we have (B.205) $= o(p^4)$. Thus,

$$\begin{aligned} &\prod_{l=1}^4 c(n,a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} (1-\mathbf{1}_E)\mathbf{1}_{\{|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|=a_1+a_2-2\}} \\ &\qquad\qquad \times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l}:l=1,2)\Big\} \\ =\ & o(1)n^{-a_1-a_2}p^{-4}n^{a_1+a_2-2}p^4 = o(n^{-2}). \end{aligned}$$

In summary, we obtain $\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - G_{a_1,a_2,1} = o(n^{-2})$.

Proof under Condition 3.4.1* Similarly to Section B.5.26, we next prove that $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| = o(n^{-2})$ and $|\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - G_{a_1,a_2,1}| = o(n^{-2})$.

Step I: $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| = o(n^{-2})$. Following the same analysis in Section B.5.26, we obtain (B.198) and (B.199). By Condition 3.4.1* (2) and (4), we have

$$
\begin{aligned}
&\sum_{1\leq j_1,j_2,j_3,j_4 j_5,j_6,j_7,j_8\leq p} (\text{B.197}) \\
= \;& O(1)\Big\{\sum_{1\leq j_1,j_2,j_3,j_4\leq p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4})^{a_1}\Big\}\Big\{\sum_{1\leq j_5,j_6,j_7,j_8\leq p}(\sigma_{j_5,j_6}\sigma_{j_7,j_8})^{a_2}\Big\}.
\end{aligned}
\tag{B.206}
$$

Note that $\sigma^2(a) = \Theta(n^{-a})\times\sum_{1\leq j_1,j_2,j_3,j_4\leq p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a$ by Lemma B.3.1, and $c(n,a) = \Theta(1)\{n^a\sigma(a)\}^{-1}$. Combining (B.199) and (B.206), we have $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2})\}^2 - G_{a_1,a_2,1}| = o(n^{-2})$.

Step II: $|\mathrm{E}(\mathbb{T}_{k,a_1,a_2}^2) - G_{a_1,a_2,1}| = o(n^{-2})$. Similarly to Section B.5.26, we have (B.201) and $\mathrm{E}\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\} \neq 0$ only when $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| \leq a_1+a_2-2$.

When $|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}| < a_1+a_2-2$, (B.203) still holds. By Condition 3.4.1* (2) and (4), similarly to (B.206), we have

$$
\begin{aligned}
&\sum_{1\leq j_1,j_2,j_3,j_4 j_5,j_6,j_7,j_8\leq p} \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\} \\
= \;& O(1)\Big\{\sum_{1\leq j_1,j_2,j_3,j_4\leq p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4})^{a_1}\Big\}\Big\{\sum_{1\leq j_5,j_6,j_7,j_8\leq p}(\sigma_{j_5,j_6}\sigma_{j_7,j_8})^{a_2}\Big\}.
\end{aligned}
$$

Note that $\sigma^2(a) = \Theta(n^{-a})\times\sum_{1\leq j_1,j_2,j_3,j_4\leq p}(\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a$ by Lemma B.3.1, and $c(n,a) = \Theta(1)\{n^a\sigma(a)\}^{-1}$. Then we have

$$
\begin{aligned}
&\prod_{l=1}^4 c(n,a_l)\sum_{\substack{1\leq j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} \mathbf{1}_{\{|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|<a_1+a_2-2\}} \\
&\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\} = o(n^{-2}).
\end{aligned}
$$

When $|\cup_{l=1}^4 \{\mathbf{i}^{(l)}\}| = a_1 + a_2 - 2$, by the construction of $\mathbf{1}_E$, we know

$$\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\Big\} \times (1 - \mathbf{1}_E) \tag{B.207}$$

satisfies that (B.207) $\neq 0$ if $(\{\mathbf{i}^{(1)}\} \cup \{\mathbf{i}^{(2)}\}) \cap (\{\mathbf{i}^{(3)}\} \cup \{\mathbf{i}^{(4)}\}) \neq \emptyset$. Then by Condition 3.4.1* (3) and (4), we know (B.207) is a linear combination of $\sum_{1\le j_1,\ldots,j_8\le p} \prod_{t=1}^{a+b} \sigma_{j_{g_{2t-1}}, j_{g_{2t}}}$ with $S_{\mathcal{G}} > 4$, where we recall that $S_{\mathcal{G}}$ is the number of distinct sets among $\{g_{2t-1}, g_{2t}\}, t = 1, \ldots, a+b$, induced by $\mathcal{G} = (g_1, \ldots, g_{2(a+b)})$. Therefore,

$$\begin{aligned}
&\prod_{l=1}^4 c(n, a_l) \sum_{\substack{1\le j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a_l-1),\, l=1,2,3,4}} (1 - \mathbf{1}_E) \times \mathbf{1}_{\{|\cup_{l=1}^4\{\mathbf{i}^{(l)}\}|=a_1+a_2-2\}} \\
&\times \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},1}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4)\Big\} \\
\le\, & C\Big\{\prod_{l=1}^4 c(n, a_l)\Big\} \times n^{a_1+a_2-2} \sum_{\mathcal{G}: S_{\mathcal{G}}>4} \Big| \sum_{1\le j_1,\ldots,j_8\le p} \prod_{t=1}^{a+b} \sigma_{j_{g_{2t-1}}, j_{g_{2t}}} \Big| = o(n^{-2}),
\end{aligned}$$

where the last equation follows by Condition 3.4.1* (4), $c(n, a) = \Theta(1)\{n^a \sigma(a)\}^{-1}$, and $\sigma^2(a) = \Theta(n^{-a}) \times \sum_{1\le j_1,j_2,j_3,j_4\le p} (\sigma_{j_1,j_3}\sigma_{j_2,j_4})^a$. In summary, we obtain $\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2}) - G_{a_1,a_2,1} = o(n^{-2})$.

Part II: $n_x \le k \le n_x + n_y$ In this section, we prove that when $n_x \le k \le n_x + n_y$, $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$. Recall the form derived in Section B.5.25. We have $\mathbb{T}_{k,a_1,a_2} = \sum_{L_1\le s_1\le a_1, L_2\le s_2\le a_2} \mathbb{T}_{k,a_1,a_2,s_1,s_2}$, where

$$\begin{aligned}
\mathbb{T}_{k,a_1,a_2,s_1,s_2} = & \sum_{\substack{1\le j_1,j_2,j_3,j_4\le p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(n_x,s_l):\, l=1,2;\\ \mathbf{w}^{(l)}\in\mathcal{P}(k-n_x-1,a_l-s_l-1):\, l=1,2}} \prod_{l=1}^2 c_2(n, a_l, s_l) \\
& \times \mathbb{M}_{\mathbf{x},\mathbf{y},2}(k - n_x, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2).
\end{aligned}$$

To prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2}) = o(n^{-2})$, it suffices to prove $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,s_1,s_2}) = o(n^{-2})$. In particular, for easy presentation, we set $a_3 = a_1$, $a_4 = a_2$ $s_3 = s_1$, and $s_4 = s_2$, and then have

$$\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,s_1,s_2})\}^2 = \sum_{\substack{1\leq j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(n_x,s_l):\, l=1,2,3,4;\\ \mathbf{w}^{(l)}\in\mathcal{P}(k-n_x-1,a_l-s_l-1):\, l=1,2,3,4}} \prod_{l=1}^{4} c_2(n,a_l,s_l)$$
$$\mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2)\Big\}$$
$$\times\, \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=3,4)\Big\}.$$

In addition, we have

$$\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,s_1,s_2}) = \sum_{\substack{1\leq j_1,j_2,j_3,j_4,j_5,j_6,j_7,j_8\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(n_x,s_l):\, l=1,2,3,4;\\ \mathbf{w}^{(l)}\in\mathcal{P}(k-n_x-1,a_l-s_l-1):\, l=1,2,3,4}} \prod_{l=1}^{4} c_2(n,a_l,s_l)$$
$$\times\, \mathrm{E}\Big\{\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)\Big\},$$

where we define

$$\mathbb{M}_{\mathbf{x},\mathbf{y},2}(k-n_x,\mathbf{i}^{(l)},j_{2l-1},j_{2l} : l=1,2,3,4)$$
$$= \prod_{l=1}^{4}\Big(\prod_{t=1}^{s_l}\mathcal{X}_{i_t^{(l)},j_{2l-1},j_{2l}}\prod_{\tilde{t}=1}^{a_l-s_l-1}\mathcal{Y}_{w_{\tilde{t}}^{(l)},j_{2l-1},j_{2l}}\Big)$$
$$\mathrm{E}(\mathcal{Y}_{k-n_x,j_1,j_2}\mathcal{Y}_{k-n_x,j_3,j_4}) \times \mathrm{E}(\mathcal{Y}_{k-n_x,j_5,j_6}\mathcal{Y}_{k-n_x,j_7,j_8}).$$

Therefore $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,s_1,s_2}) = \mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,s_1,s_2}) - \{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,s_1,s_2})\}^2$ is derived. We note that the form of $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,s_1,s_2})$ is very similar to the $\mathrm{var}(\mathbb{T}_{k,a_1,a_2})$ in Section B.5.26. In particular, we can write $\mathcal{Z}_{i,j_1,j_2} = \mathcal{X}_{i,j_1,j_2}$ if $i \leq n_x$ and $\mathcal{Z}_{i,j_1,j_2} = \mathcal{Y}_{i-n_x,j_1,j_2}$ if $i > n_x$. Then we let $\mathbf{q}^{(l)} = (\mathbf{i}^{(l)}, \tilde{\mathbf{w}}^{(l)})$ to be a joint index tuple of $\mathbf{i}^{(l)}$ and $\mathbf{w}^{(l)}$, where $\tilde{\mathbf{w}}^{(l)}$ is transformed from $\mathbf{w}^{(l)}$ by adding each index with n_x. Also let $\mathbf{1}_{\tilde{E}}$ be an indicator function of the event that $(\{\mathbf{q}^{(1)}\} \cup \{\mathbf{q}^{(2)}\}) \cap (\{\mathbf{q}^{(3)}\} \cup \{\mathbf{q}^{(4)}\}) = \emptyset$. Then define

$$
\begin{aligned}
G_{a_1,a_2,2} = \prod_{l=1}^{4} c(n, a_l) & \sum_{\substack{1 \le j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8 \le p; \\ \mathbf{i}^{(l)} \in \mathcal{P}(n_x, s_l):\, l=1,2,3,4; \\ \mathbf{w}^{(l)} \in \mathcal{P}(k-n_x-1, a_l-s_l-1):\, l=1,2,3,4}} \times \mathbf{1}_{\tilde{E}} \\
& \times \mathrm{E}\Big\{ \mathbb{M}_{\mathbf{x},\mathbf{y},2}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4) \Big\}.
\end{aligned}
$$

Similarly to Section B.5.26, we also note that

$$
\begin{aligned}
& \mathrm{E}\Big\{ \mathbb{M}_{\mathbf{x},\mathbf{y},2}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2, 3, 4) \Big\} \times \mathbf{1}_{\tilde{E}} \\
= & \mathrm{E}\Big\{ \mathbb{M}_{\mathbf{x},\mathbf{y},2}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 1, 2) \Big\} \\
& \times \mathrm{E}\Big\{ \mathbb{M}_{\mathbf{x},\mathbf{y},2}(k, \mathbf{i}^{(l)}, j_{2l-1}, j_{2l} : l = 3, 4) \Big\} \times \mathbf{1}_{\tilde{E}}.
\end{aligned}
$$

Given Conditions 3.4.1 and 3.4.1*, we know that similarly to Section B.5.26, we can show $|\{\mathrm{E}(\mathbb{T}_{k,a_1,a_2,s_1,s_2})\}^2 - G_{a_1,a_2,2}| = o(n^{-2})$ and $|\mathrm{E}(\mathbb{T}^2_{k,a_1,a_2,s_1,s_2}) - G_{a_1,a_2,2}| = o(n^{-2})$ respectively. Finally we obtain $\mathrm{var}(\mathbb{T}_{k,a_1,a_2,s_1,s_2}) = o(n^{-2})$. The proof is very similar and the details is thus skipped.

B.5.27 Proof of Lemma B.3.4

Recall the form of $D_{n,k}$ derived in Section B.5.25:

$$
\sum_{k=1}^{n} \mathrm{E}(D_{n,k}^4) = \sum_{k=1}^{n} \sum_{1 \le r_1, r_2, r_3, r_4 \le m} \prod_{l=1}^{4} t_{r_l} \times \mathrm{E}\Big(\prod_{l=1}^{4} A_{n,k,a_{r_l}} \Big).
$$

To prove Lemma B.3.4, it suffices to show that for given $1 \le k \le n$ and $1 \le r_1, r_2, r_3, r_4 \le m$, we have $\mathrm{E}(\prod_{l=1}^{4} A_{n,k,a_{r_l}}) = o(n^{-1})$. In addition, by the Cauchy-Schwarz inequality, it suffices to show $\mathrm{E}(A_{n,k,a}^4) = o(n^{-1})$ for each given finite a.

Part I: $1 \leq k \leq n_x$ We consider without loss of generality that $k \geq a$ and

$$\mathrm{E}\Big(\prod_{l=1}^{4} A_{n,k,a}^{4}\Big) = c^4(n,a) \sum_{\substack{1\leq j_1,\ldots,j_8\leq p;\\ \mathbf{i}^{(l)}\in\mathcal{P}(k-1,a-1),\, l=1,\ldots,4}} \mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t_l=1}^{a-1} \mathcal{X}_{i_{t_l}^{(l)},j_{2l-1},j_{2l}}\Big) \times \mathrm{E}\Big(\prod_{l=1}^{4}\mathcal{X}_{j_{2l-1},j_{2l}}\Big).$$

As $\mathrm{E}(\mathcal{X}_{j_1,j_2}) = 0$ under H_0, we know

$$\mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t_l=1}^{a-1} \mathcal{X}_{i_{t_l}^{(l)},j_{2l-1},j_{2l}}\Big) \neq 0$$

only when $|\cup_{l=1}^{4}\{\mathbf{i}^{(l)}\}| \leq 2(a-1)$. Note that $c(n,a) = \Theta(1)\{n^a\sigma(a)\}^{-1}$. To finish the proof, it suffices to show that for given $(\mathbf{i}^{(1)},\mathbf{i}^{(2)},\mathbf{i}^{(3)},\mathbf{i}^{(4)})$, we have

$$\sum_{1\leq j_1,\ldots,j_8\leq p} \mathrm{E}\Big(\prod_{l=1}^{4}\prod_{t_l=1}^{a-1} \mathcal{X}_{i_{t_l}^{(l)},j_{2l-1},j_{2l}}\Big)\mathrm{E}\Big(\prod_{l=1}^{4}\mathcal{X}_{j_{2l-1},j_{2l}}\Big) = O(n^{2a})\sigma^4(a) \tag{B.208}$$

We next prove (B.208) under Conditions 3.4.1 and 3.4.1* in the following Sections B.5.27 and B.5.27, respectively.

Under Condition 3.4.1 Recall that $\mathcal{X}_{i,j_1,j_2} = x_{i,j_1}x_{i,j_2} - \sigma_{j_1,j_2}$. By the symmetricity of the j indexes, we have

$$\sum_{1\leq j_1,\ldots,j_8\leq p}\Big|\mathrm{E}\Big(\prod_{l=1}^{4}\mathcal{X}_{j_{2l-1},j_{2l}}\Big)\Big| \leq C \sum_{1\leq j_1,\ldots,j_8\leq p}\bigg\{\Big|\mathrm{E}\Big(\prod_{l=1}^{8}x_{1,j_l}\Big)\Big| + \Big|\mathrm{E}\Big(\prod_{l=1}^{6}x_{1,j_l}\Big)\sigma_{j_7,j_8}\Big| + \Big|\mathrm{E}\Big(\prod_{l=1}^{4}x_{1,j_l}\Big)\sigma_{j_5,j_6}\sigma_{j_7,j_8}\Big| + \Big|\prod_{l=1}^{4}\sigma_{j_{2l-1},\,j_{2l}}\Big|\bigg\}.$$

Under Condition 3.4.1 with the mixing-type assumption, following analyses similar to those in Sections B.5.6.1 and B.5.7.1, we know that $\sum_{1\leq j_1,\ldots,j_8\leq p}|\mathrm{E}(\prod_{l=1}^{8}x_{1,j_l})|$, $\sum_{1\leq j_1,\ldots,j_8\leq p}|\mathrm{E}(\prod_{l=1}^{6}x_{1,j_l})\sigma_{j_7,j_8}|$, $\sum_{1\leq j_1,\ldots,j_8\leq p}|\mathrm{E}(\prod_{l=1}^{4}x_{1,j_l})\sigma_{j_5,j_6}\sigma_{j_7,j_8}|$ and

$\sum_{1\le j_1,\ldots,j_8\le p}|\prod_{l=1}^4 \sigma_{j_{2l-1},j_{2l}}|$ are all $O(p^4)$. It follows that

$$\sum_{1\le j_1,\ldots,j_8\le p}\Big|\mathrm{E}\Big(\prod_{l=1}^4 \mathcal{X}_{j_{2l-1},j_{2l}}\Big)\Big| = O(p^4). \tag{B.209}$$

Recall that Lemma B.3.1 shows that $\sigma^2(a)=\Theta(p^2n^{-a})$. By (B.209) and Condition 3.4.1 (2), we have (B.208) holds and $\mathrm{E}(A_{n,k,a}^4)=o(n^{-1})$.

Under Condition 3.4.1* By Condition 3.4.1* (3), we know that the moment $\mathrm{E}(\prod_{l=1}^4\prod_{t_l=1}^{a-1}\mathcal{X}_{i_{t_l}^{(l)},j_{2l-1},j_{2l}})\times\mathrm{E}(\prod_{l=1}^4\mathcal{X}_{j_{2l-1},j_{2l}})$ is a linear combination of $\mathrm{E}(\prod_{t=1}^{4a}\sigma_{j_{g_{2t-1}},j_{g_{2t}}})$, where $\mathcal{G}=(g_1,\ldots,g_{8a})\in\{1,\ldots,8\}^{8a}$ satisfies that $g_{2t-1}\neq g_{2t}$ for $t=1,\ldots,4a$ and the number of g's equal to m is a for each $m\in\{1,\ldots,8\}$. By Condition 3.4.1* (4), for given $\mathcal{G}$ satisfying the constraints, $\sum_{1\le j_1,\ldots,j_8\le p}\sigma_{j_{g_{2t-1}},j_{g_{2t}}}=O(1)\sum_{1\le j_1,\ldots,j_8\le p}(\sigma_{j_1,j_2}\times\sigma_{j_3,j_4}\sigma_{j_5,j_6}\sigma_{j_7,j_8})^a$. Then we have

$$\begin{aligned}&\sum_{1\le j_1,\ldots,j_8\le p}\mathrm{E}\Big(\prod_{l=1}^4\prod_{t_l=1}^{a-1}\mathcal{X}_{i_{t_l}^{(l)},j_{2l-1},j_{2l}}\Big)\times\mathrm{E}(\prod_{l=1}^4\mathcal{X}_{j_{2l-1},j_{2l}})\\ &=O(1)\sum_{1\le j_1,\ldots,j_8\le p}(\sigma_{j_1,j_2}\sigma_{j_3,j_4}\sigma_{j_5,j_6}\sigma_{j_7,j_8})^a=O(1)\Big(\sum_{1\le j_1,j_2\le p}\sigma_{j_1,j_2}^a\Big)^4.\end{aligned}$$

Recall that Lemma B.3.1 shows that $\sigma^2(a)=\Theta(n^{-a})(\sum_{1\le j_1,j_2\le p}\sigma_{j_1,j_2}^a)^2$. Therefore, (B.208) is obtained and Lemma B.3.4 is proved.

Part II: $n_x+1\le k\le n_x+n_y$ Section B.5.25 derives that $A_{n,k,a}=\sum_{s=L_k}^{a-1}A_{n,k,a,s}$, where

$$A_{n,k,a,s}=\sum_{\substack{1\le j_1,j_2\le p;\\ \mathbf{i}\in\mathcal{P}(n_x,s);\\ \mathbf{w}\in\mathcal{P}(k-n_x-1,a-s-1)}} c_2(n,a,s)\mathcal{Y}_{k-n_x,j_1,j_2}\prod_{t=1}^s\mathcal{X}_{i_t,j_1,j_2}\prod_{\tilde{t}=1}^{a-s-1}\mathcal{Y}_{w_{\tilde{t}},j_1,j_2}.$$

Similarly to Section B.5.27, it suffices to show that for given finite integers a and s, $\mathrm{E}(A_{n,k,a,s}^4)=o(n^{-1})$. Following the arguments in Section B.5.26, we know $A_{n,k,a,s}$

takes a similar form to $A_{n,k,a}$ in Section B.5.27. Therefore the proof in Section B.5.27 can be applied similarly to show $\mathrm{E}(A^4_{n,k,a,s}) = o(n^{-1})$ in this section. The proof will be very similar and the details are thus skipped.

B.5.28 Proof of Lemma B.3.5

In this section, to prove Lemma B.3.5, we study $\mathrm{var}(T_{D,a,1})$, $\mathrm{var}(T_{D,a,2})$ and $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$ respectively.

Part I: $\mathrm{var}(T_{D,a,1})$ We first derive $\mathrm{var}(T_{D,a,1})$. Note that $T_{D,a,1}$ is a summation over j indexes in $\mathbb{J}_0$, and $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2}$ for $j_1, j_2 \in \mathbb{J}_0$. Following the arguments in Section B.5.23, similarly to (B.181), we have

$$\mathrm{var}(T_{D,a,1}) \simeq \sum_{1 \le j_1,j_2,j_3,j_4 \in \mathbb{J}_0} a!(\mathbf{X}_{j_1,j_2,j_3,j_4}/n_x + \mathbf{Y}_{j_1,j_2,j_3,j_4}/n_y)^a.$$

By Condition B.3.1 (3), (B.189) still holds. Then by Condition B.3.2 and the symmetricity of j indexes,

$$\mathrm{var}(T_{D,a,1}) \simeq C_{\kappa,a} \sum_{1 \le j_1,j_2,j_3,j_4 \in \mathbb{J}_0} a!\sigma^a_{j_1,j_2}\sigma^a_{j_3,j_4}, \tag{B.210}$$

where $C_{\kappa,a} = \{(\kappa_x - 1)/n_x + (\kappa_y - 1)/n_y\}^a + 2(\kappa_x/n_x + \kappa_y/n_y)^a$, and $\mathrm{var}(T_{D,a,1})$ is of order $\Theta(n^{-a}\mathbb{V}^{1/2}_{a,a,0,0})$ with $\mathbb{V}^{1/2}_{a,a,0,0} = \sum_{j_1,\ldots,j_4 \in \mathbb{J}_0}(\sigma_{x,j_1,j_2}\sigma_{x,j_3,j_4})^a$ defined on Page 278.

Part II: $\mathrm{var}(T_{D,a,2})$ We show $\mathrm{var}(T_{D,a,2}) = o(1)\mathrm{var}(T_{D,a,1})$. Particularly,

$$T_{D,a,2} = \sum_{(j_1,j_2) \in J_{0,D}} \frac{1}{P^{n_x}_a P^{n_y}_a} \sum_{\substack{\mathbf{i} \in \mathcal{P}(n_x,a),\\ \mathbf{w} \in \mathcal{P}(n_y,a)}} \prod_{t=1}^{a} (\mathcal{X}_{i_t,j_1,j_2} - \mathcal{Y}_{w_t,j_1,j_2}),$$

where we redefine $\mathcal{X}_{i,j_1,j_2} = x_{i,j_1}x_{i,j_2} - \sigma_{y,j_1,j_2}$ and $\mathcal{Y}_{i,j_1,j_2} = y_{i,j_1}y_{i,j_2} - \sigma_{y,j_1,j_2}$. Moreover, we define

$$G_{D,a} = \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} (P_a^{n_x} P_a^{n_y})^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a),\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} \mathbf{1}_{\{\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}=\emptyset\}} (D_{j_1,j_2}D_{j_3,j_4})^a.$$

To prove $\mathrm{var}(T_{D,a,2}) = \mathrm{E}(T_{D,a,2}^2) - \{\mathrm{E}(T_{D,a,2})\}^2$ is $o(1)\mathrm{var}(T_{D,a,1})$, we next show $|\mathrm{E}(T_{D,a,2}^2) - G_{D,a}|$ and $|\{\mathrm{E}(T_{D,a,2})\}^2 - G_{D,a}|$ are both $o(1)\mathrm{var}(T_{D,a,1})$.

Note that $\mathrm{E}(\mathcal{X}_{i,j_1,j_2}) = D_{j_1,j_2}$ and $\mathrm{E}(\mathcal{Y}_{i,j_1,j_2}) = 0$. We have

$$\{\mathrm{E}(T_{D,a,2})\}^2 = \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} (P_a^{n_x} P_a^{n_y})^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a),\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} (D_{j_1,j_2}D_{j_3,j_4})^a.$$

Then

$$\begin{aligned}
&|\{\mathrm{E}(T_{D,a,2})\}^2 - G_{D,a}| \\
\leq& \Bigg| \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} (P_a^{n_x} P_a^{n_y})^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a),\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} \mathbf{1}_{\{\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}\neq\emptyset\}} (D_{j_1,j_2}D_{j_3,j_4})^a \Bigg| \\
\leq& Cn^{-1} \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |D_{j_1,j_2}D_{j_3,j_4}|^a,
\end{aligned}$$

where we use $\sum_{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a),\mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)} \mathbf{1}_{\{\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}\neq\emptyset\}} = O(n^{4a-1})$. In addition,

$$\begin{aligned}
&|\mathrm{E}(T_{D,a,2}^2) - G_{D,a}| \\
\leq& C \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} (P_a^{n_x} P_a^{n_y})^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,a),\\ \mathbf{w},\tilde{\mathbf{w}}\in\mathcal{P}(n_y,a)}} \\
&\Bigg(\mathbf{1}_{\{\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}=\emptyset\}} \Bigg| \mathrm{E}\Big\{ \prod_{t=1}^{a} (\mathcal{X}_{i_t,j_1,j_2} - \mathcal{Y}_{w_t,j_1,j_2})(\mathcal{X}_{\tilde{i}_t,j_3,j_4} - \mathcal{Y}_{\tilde{w}_t,j_3,j_4}) \Big\} - (D_{j_1,j_2}D_{j_3,j_4})^a \Bigg| \\
&+ \mathbf{1}_{\{\{\mathbf{i}\}\cap\{\tilde{\mathbf{i}}\}\neq\emptyset\}} \Bigg| \mathrm{E}\Big\{ \prod_{t=1}^{a} (\mathcal{X}_{i_t,j_1,j_2} - \mathcal{Y}_{w_t,j_1,j_2})(\mathcal{X}_{\tilde{i}_t,j_3,j_4} - \mathcal{Y}_{\tilde{w}_t,j_3,j_4}) \Big\} \Bigg| \Bigg).
\end{aligned}$$

We redefine $\mathbf{X}_{j_1,j_2,j_3,j_4} = \mathrm{E}(\mathcal{X}_{i,j_1,j_2}\mathcal{X}_{i,j_3,j_4})$ and $\mathbf{Y}_{j_1,j_2,j_3,j_4} = \mathrm{E}(\mathcal{Y}_{i,j_1,j_2}\mathcal{Y}_{i,j_3,j_4})$. Then

$$\begin{aligned}
&|\mathrm{E}(T_{D,a,2}^2) - G_{D,a}| \\
\leq\, & C \sum_{1\leq m_1+m_2\leq a} n^{-m_1-m_2} \\
&\times \sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} \left|\mathbf{X}^{m_1}_{j_1,j_2,j_3,j_4}\mathbf{Y}^{m_2}_{j_1,j_2,j_3,j_4}(D_{j_1,j_2}D_{j_3,j_4})^{a-m_1-m_2}\right|.
\end{aligned}$$

Note that $\mathbf{Y}_{j_1,j_2,j_3,j_4} = \sigma_{y,j_1,j_3}\sigma_{y,j_2,j_4} + \sigma_{y,j_1,j_4}\sigma_{y,j_2,j_3}$ and $\sigma_{y,j_1,j_2} = \sigma_{x,j_1,j_2} - D_{j_1,j_2}$. By Conditions B.3.1 and B.3.2, Hölder's inequality, and definitions in (B.26), we have

$$\mathrm{var}(T_{D,a,2}) \leq C \max_{\substack{\mathcal{H}\in\mathbb{H},\\ t=1,2}} \Big\{ \sum_{m=1}^{a} (n^{-a}\mathbb{V}_{a,\mathcal{H},x,t})^{m/a}(\mathbb{V}_{a,\mathcal{H},D,3})^{1-m/a},\, n^{-1}\mathbb{V}_{a,\mathcal{H},D,3}\Big\}.$$

Therefore by Condition B.3.2 and (B.210), $\mathrm{var}(T_{D,a,2}) = o(1)n^{-a}\mathbb{V}^{1/2}_{a,a,0,0} = o(1)\mathrm{var}(T_{D,a,1})$.

Part III: $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\}$ Last, we prove $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}(T_{D,a,1})$. Similarly to Section B.5.23, we write $\tilde{\mathcal{U}}^*(a) = \sum_{c=0}^{a}\sum_{b_1=0}^{c}\sum_{b_2=0}^{a-c} C_{a,c,b_1,b_2}\times T_{b_1,b_2,c}\mathbf{1}_{b_1+b_2\leq a-1}$, where $T_{b_1,b_2,c}$ is defined in (B.185). For finite a, to prove $\mathrm{var}\{\tilde{\mathcal{U}}^*(a)\} = o(1)\mathrm{var}(T_{D,a,1})$, it suffices to prove $\mathrm{var}(T_{b_1,b_2,c}) = o(1)\mathrm{var}(T_{D,a,1})$ for $0 \leq c \leq a$ and $b_1 + b_2 \leq a-1$. As $\mathrm{E}(\mathcal{Y}_{i,j_1,j_2}) = 0$ and $\mathrm{E}(\mathbf{x}) = \mathrm{E}(\mathbf{y}) = 0$, we know that if $b_1 + b_2 \leq a-1$, $\mathrm{E}(T_{b_1,b_2,c}) = 0$. Then $\mathrm{var}(T_{b_1,b_2,c}) = \mathrm{E}(T^2_{b_1,b_2,c})$, which takes a similar form to (B.186). Specifically, we can write $\mathrm{var}(T_{b_1,b_2,c}) = \mathrm{var}(T_{b_1,b_2,c})_{(1)} + \mathrm{var}(T_{b_1,b_2,c})_{(2)}$, where

$$\mathrm{var}(T_{b_1,b_2,c})_{(1)} = \sum_{j_1,j_2,\tilde{j}_1,\tilde{j}_2\in\mathbb{J}_0} (P^{n_x}_{2c-b_1}P^{n_y}_{2(a-c)-b_2})^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,2c-b_1);\\ \mathbf{w}\,\tilde{\mathbf{w}}\in\mathcal{P}(n_y,2(a-c)-b_2)}} \mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2),$$

and

$$\mathrm{var}(T_{b_1,b_2,c})_{(2)} = \sum_{\substack{(j_1,j_2),\\(\tilde{j}_1,\tilde{j}_2)\in J_{0,D}}} (P^{n_x}_{2c-b_1}P^{n_y}_{2(a-c)-b_2})^{-2} \sum_{\substack{\mathbf{i},\tilde{\mathbf{i}}\in\mathcal{P}(n_x,2c-b_1);\\ \mathbf{w}\,\tilde{\mathbf{w}}\in\mathcal{P}(n_y,2(a-c)-b_2)}} \mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2),$$

and $\mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2)$ is defined same as in (B.186).

Note that $\mathrm{var}(T_{b_1,b_2,c})_{(1)}$ is a summation over j indexes in $\mathbb{J}_0$, and $\sigma_{x,j_1,j_2} = \sigma_{y,j_1,j_2}$ for $j_1, j_2 \in \mathbb{J}_0$. Therefore the arguments under H_0 in Section B.5.23 can be applied similarly to $\mathrm{var}(T_{b_1,b_2,c})_{(1)}$. Then we have $\mathrm{var}(T_{b_1,b_2,c})_{(1)} = o(n^{-a})(\sum_{j_1,j_2\in\mathbb{J}_0}\sigma^a_{j_1,j_2})^2$ which is $o(1)\mathrm{var}(T_{D,a,1})$. We next consider $\mathrm{var}(T_{b_1,b_2,c})_{(2)}$. As $\mathrm{E}(\mathcal{Y}_{i,j_1,j_2}) = 0$ and $\mathrm{E}(\mathbf{x}) = \mathrm{E}(\mathbf{y}) = 0$, by the definition in (B.186), we know $\mathrm{E}\{\mathbb{T}(\mathbf{i},\tilde{\mathbf{i}},\mathbf{w},\tilde{\mathbf{w}},j_1,j_2,\tilde{j}_1,\tilde{j}_2)\} \neq 0$ only when $\{i_{b_1+1},\dots,i_{2c-b_1}\} = \{\tilde{i}_{b_1+1},\dots,\tilde{i}_{2c-b_1}\}$ and $\{\mathbf{w}\} = \{\tilde{\mathbf{w}}\}$. Let $m_0 = b_1 - |\{i_1,\dots,i_{b_1}\}\cap\{\tilde{i}_1,\dots,\tilde{i}_{b_1}\}|$. By Condition B.3.1 (3) and Hölder's inequality,

$$\begin{aligned}
&\mathrm{var}(T_{b_1,b_2,c})_{(2)}\\
&\leq Cn_x^{-(c-b_1)}n_y^{-(a-c-b_2)} \max_{\substack{\mathcal{H}\in\mathbb{H}_0,\\0\leq m_0\leq b_1}} \Bigg\{\Bigg(n_y^{-a}\sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |\sigma_{y,j_{h_1},j_{h_2}}\sigma_{y,j_{h_3},j_{h_4}}|^a\Bigg)^{\frac{a-c}{a}}\\
&\quad\times\Bigg(n_x^{-a}\sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |\sigma_{x,j_{h_1},j_{h_2}}\sigma_{x,j_{h_3},j_{h_4}}|^a\Bigg)^{\frac{c-m_0}{a}}\\
&\quad\times\Bigg(\sum_{(j_1,j_2),(j_3,j_4)\in J_{0,D}} |D_{j_{h_1},j_{h_2}}D_{j_{h_3},j_{h_4}}|^a\Bigg)^{\frac{m_0}{a}}\Bigg\}\\
&\leq Cn^{-(a-b_1-b_2)}\max_{\mathcal{H}\in\mathbb{H}_0,t=1,2}\{n^{-a}\mathbb{V}_{a,\mathcal{H},x,t},\mathbb{V}_{a,\mathcal{H},D,3}\},
\end{aligned}$$

where the last inequality uses $\sigma_{y,j_1,j_2} = \sigma_{x,j_1,j_2} - D_{j_1,j_2}$. Since $b_1 + b_2 \leq a-1$, we obtain $\mathrm{var}(T_{b_1,b_2,c})_{(2)} \leq Cn^{-1}\max_{\mathcal{H}\in\mathbb{H}_0;t=1,2}\{n^{-a}\mathbb{V}_{a,\mathcal{H},x,t},\mathbb{V}_{a,\mathcal{H},D,3}\}$. By Condition B.3.2 and (B.210), we know $\mathrm{var}(T_{b_1,b_2,c})_{(2)} = o(1)\mathrm{var}(T_{D,a,1})$.

B.6 Computation of U-Statistics in Chapter III

B.6.1 Closed-Form Formulae for (3.18)

Note that $\mathcal{U}_l(a) = U_l^{\mathbf{1}_a}$ by the definitions in (3.17), and for different l's, the computation methods of $U_l^{\mathbf{1}_a}$'s are the same. Therefore in the following, for simplicity, we give the formulae of $U_l^{\mathbf{1}_a}$ without the subscript l:

$$
\begin{aligned}
U^{\mathbf{1}_1} =& V^{(1)},\\
U^{\mathbf{1}_2} =& V^{(1,1)} - V^{(2)},\\
U^{\mathbf{1}_3} =& V^{\mathbf{1}_3} - 3V^{(2,1)} + 2V^{(3)},\\
U^{\mathbf{1}_4} =& V^{\mathbf{1}_4} - 6V^{(2,1,1)} + 8V^{(3,1)} + 3V^{(2,2)} - 6V^{(4)},\\
U^{\mathbf{1}_5} =& V^{\mathbf{1}_5} - 10V^{(2,\mathbf{1}_3)} + 20V^{(3,\mathbf{1}_2)} + 15V^{(2,2,1)} - 30V^{(4,1)} - 20V^{(2,3)} + 24V^{(5)},\\
U^{\mathbf{1}_6} =& V^{\mathbf{1}_6} - 15V^{(\mathbf{1}_4,2)} + 40V^{(3,\mathbf{1}_3)} + 45V^{(1,1,2,2)} - 90V^{(1,1,4)} - 120V^{(1,2,3)}\\
& + 144V^{(1,5)} - 15V^{(2,2,2)} + 90V^{(2,4)} + 40V^{(3,3)} - 120V^{(6)},
\end{aligned}
$$

where $U^{\mathbf{1}_a}$ and $V^{(t_1,\ldots,t_k)}$ are defined as in (3.17).

B.6.2 Computation with unknown mean

In this section, we provide the details of the computation of $\mathcal{U}(a)$ when $\mathrm{E}(x_{i,j})$ is unknown. We note that $\mathcal{U}(a)$ is some linear combination of

$$\sum_{1\le i_1\neq\ldots\neq i_k\le n}\prod_{t=1}^{k} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}, \tag{B.211}$$

where $a \le k \le 2a$, $r_{t,1}, r_{t,2} \ge 0$ and $r_{t,1} + r_{t,2} \ge 1$. A direct calculation of (B.211) has computational cost $O(n^k)$, which is large when k is large. But following the discussion in Section 3.2.3, we can similarly reduce the computational cost of (B.211) to order

$O(n)$ with an iterative method. In particular, we note that

$$
\begin{aligned}
&\sum_{1\le i_1\ne\ldots\ne i_k\le n}\prod_{t=1}^{k} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}} \\
=&\Big(\sum_{1\le i_1\ne\ldots\ne i_{k-1}\le n}\prod_{t=1}^{k-1} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}\Big)\Big(\sum_{i=1}^{n} x_{i,j_1}^{r_{k,1}} x_{i,j_2}^{r_{k,2}}\Big) \\
&-\sum_{m=1}^{k-1}\sum_{1\le i_1\ne\ldots\ne i_{k-1}\le n}\Big(\prod_{t=1}^{k-1} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}\Big) x_{i_m,j_1}^{r_k,1} x_{i_m,j_2}^{r_k,2}.
\end{aligned}
\tag{B.212}
$$

Suppose for any given integers $\{(r_{t,1}, r_{t,2}) : t = 1, \ldots, k-1\}$, we can compute $\sum_{1\le i_1\ne\ldots\ne i_{k-1}\le n}\prod_{t=1}^{k-1} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}$ with cost $O(n)$. Then by the relationship in (B.212), we can obtain (B.211) with cost $O(n)$ iteratively.

We then illustrate the iterative method with some examples. When $k = 1$, for any given $(r_{1,1}, r_{1,2})$, we know $\sum_{i=1}^{n} x_{i,j_1}^{r_{1,1}} x_{i,j_2}^{r_{1,2}}$ can be computed with cost $O(n)$. When $k = 2$, by (B.212), we have $\sum_{1\le i_1\ne i_2\le n}\prod_{t=1}^{2} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}} = (\sum_{i=1}^{n} x_{i,j_1}^{r_{1,1}} x_{i,j_2}^{r_{1,2}})(\sum_{i=1}^{n} x_{i,j_1}^{r_{2,1}} x_{i,j_2}^{r_{2,2}}) - \sum_{i=1}^{n} x_{i,j_1}^{r_{1,1}+r_{2,1}} x_{i,j_2}^{r_{1,2}+r_{2,2}}$, which can be computed with cost $O(n)$. For a general $k \ge 1$ and any given integers $\{(r_{t,1}, r_{t,2}) : t = 1, \ldots, k-1\}$, suppose we can compute $\sum_{1\le i_1\ne\ldots\ne i_{k-1}\le n}\prod_{t=1}^{k-1} x_{i_t,j_1}^{r_{t,1}} x_{i_t,j_2}^{r_{t,2}}$ with cost $O(n)$. Then by (B.212), we can obtain (B.211) with computational cost $O(n)$.

Given the iterative method discussed above, we can compute $\mathcal{U}(a)$ with cost $O(p^2 n)$. For example, we can write $\mathcal{U}(1)$ as

$$
\sum_{1\le j_1\ne j_2\le p}\Big\{n^{-1}\sum_{i=1}^{n} x_{i,j_1}x_{i,j_2} - (P_2^n)^{-1}\Big(\sum_{i_1=1}^{n} x_{i_1,j_1}\sum_{i_2=1}^{n} x_{i_2,j_2} - \sum_{i=1}^{n} x_{i,j_1}x_{i,j_2}\Big)\Big\}.
$$

For $a = 2$, similar analysis holds. Note that

$$
\mathcal{U}(2) = \sum_{1\le j_1\ne j_2\le p}\Big\{(P_2^n)^{-1}\mathcal{U}_1(2) - 2(P_3^n)^{-1}\mathcal{U}_2(2) + (P_4^n)^{-1}\mathcal{U}_3(2)\Big\},
$$

where

$$\mathcal{U}_1(2) = \sum_{1\le i_1\ne i_2\le n} \prod_{t=1}^{2} x_{i_t,j_1} x_{i_t,j_2},$$

$$\mathcal{U}_2(2) = \sum_{1\le i_1\ne i_2\ne i_3\le n} (x_{i_1,j_1} x_{i_1,j_2})(x_{i_2,j_1})(x_{i_3,j_2}),$$

$$\mathcal{U}_3(2) = \sum_{1\le i_1\ne i_2\ne i_3\ne i_4\le n} \prod_{t=1}^{2} x_{i_t,j_1} \prod_{t=3}^{4} x_{i_t,j_2}.$$

We then find that $\mathcal{U}_1(2)$,$\mathcal{U}_2(2)$ and $\mathcal{U}_3(2)$ can be computed with cost $O(n)$ using the following formulae.

$$\mathcal{U}_1(2) = \Big(\sum_{i=1}^{n} x_{i,j_1} x_{i,j_2}\Big)^2 - \sum_{i=1}^{n} (x_{i,j_1} x_{i,j_2})^2.$$

$$\begin{aligned}\mathcal{U}_2(2) =& \Big(\sum_{i=1}^{n} x_{i,j_1} x_{i,j_2}\Big)\Big(\sum_{1\le i_1\ne i_2\le n} x_{i,j_1} x_{i,j_2}\Big)\\ &- \sum_{1\le i_1\ne i_2\le n} (x_{i_1,j_1}^2 x_{i_1,j_2}) x_{i_2,j_2} - \sum_{1\le i_1\ne i_2\le n} (x_{i_1,j_1} x_{i_1,j_2}^2) x_{i_2,j_1},\end{aligned}$$

where we use $\sum_{1\le i_1\ne i_2\le n} x_{i,j_1} x_{i,j_2} = (\sum_{i=1}^{n} x_{i,j_1})(\sum_{i=1}^{n} x_{i,j_2}) - \sum_{i=1}^{n} x_{i,j_1} x_{i,j_2}$, and $\sum_{1\le i_1\ne i_2\le n} (x_{i_1,j_1}^2 x_{i_1,j_2}) x_{i_2,j_2} = (\sum_{i=1}^{n} x_{i,j_1}^2 x_{i,j_2})(\sum_{i=1}^{n} x_{i,j_2}) - \sum_{i=1}^{n} x_{i,j_1}^2 x_{i,j_2}^2$.

$$\mathcal{U}_3(2) = \Big(\sum_{1\le i_1\ne i_2\le n} x_{i_1,j_1} x_{i_2,j_1}\Big)\Big(\sum_{1\le i_3\ne i_4\le n} x_{i_3,j_2} x_{i_4,j_2}\Big) - 2\mathcal{U}_1(2) - 4\mathcal{U}_3(2),$$

where we use $\sum_{1\le i_1\ne i_2\le n} x_{i_1,k} x_{i_2,k} = (\sum_{i=1}^{n} x_{i,k})^2 - \sum_{i=1}^{n} x_{i,k}^2$ for $k = j_1, j_2$.

When $a \ge 3$, the similar iterative method can be applied. But the closed form for computation might be hard to derive directly. Alternatively, we introduce a simplified form of U-statistics: $\mathcal{U}_c(a) = (P_a^n)^{-1} \sum_{1\le i_1\ne\ldots\ne i_a\le n} \sum_{1\le j_1\ne j_2\le p} \prod_{t=1}^{a} (x_{i_t,j_1} - \bar{x}_{j_1})(x_{i_t,j_2} - \bar{x}_{j_2})$. We note that $\mathcal{U}_c(a)$ takes a similar form to $\tilde{\mathcal{U}}(a)$ in (3.5), but replacing each observation $x_{i,j}$ with the centered correspondence $x_{i,j} - \bar{x}_j$. Therefore, $\mathcal{U}_c(a)$ can be computed with cost $O(n)$ using Algorithm III.1, if we set $s_{i,l} = (x_{i,j_1} - \bar{x}_{j_1})(x_{i,j_2} - \bar{x}_{j_2})$ in Algorithm III.1 for $l \in \{(j_1, j_2) : 1 \leq j_1 \neq j_2 \leq p\}$. We then show that we can substitute $\mathcal{U}(a)$ with $\mathcal{U}_c(a)$ when $a \geq 3$ in computation under certain conditions.

Proposition B.6.1. *Under the Conditions of Theorem 3.2.4, consider $a \geq 3$. If a is odd, $p = o(n^{1+a/2})$; if a is even, $p = o(n^{a/2})$. Then $\{\mathcal{U}(a) - \mathcal{U}_c(a)\}/\sigma(a) \xrightarrow{P} 0$.*

The proof of Proposition B.6.1 can be found in Section C.1.3 of He et al. (2021e). It implies that the results in Theorem 3.2.4 sill hold by replacing $\mathcal{U}(a)$ with $\mathcal{U}_c(a)$. As discussed above, we recommend including U-statistics of orders $\{1, 2, 3, \ldots, 6, \infty\}$ in the adaptive testing procedure. Then Proposition B.6.1 requires that $p = o(n^2)$, which suits a wide range of applications. Combining Theorem 3.2.4 and Proposition B.6.1, we can conduct the test with quick computation of cost $O(p^2 n)$.

On the other hand, we can conduct the test more generally without Condition 3.2.4 and the requirement $p = o(n^2)$. Specifically, we compute $\tilde{\mathcal{U}}(a)$ in (3.5) with cost $O(p^2 n)$. Then $[\tilde{\mathcal{U}}(a) - \mathrm{E}\{\tilde{\mathcal{U}}(a)\}]/\sqrt{\mathrm{var}\{\tilde{\mathcal{U}}(a)\}} \xrightarrow{D} \mathcal{N}(0,1)$ by Lemma B.1.1 and Theorem 3.2.4. To test H_0 in (3.2), it suffices to estimate $\mathrm{E}\{\tilde{\mathcal{U}}(a)\}$ and $\mathrm{var}\{\tilde{\mathcal{U}}(a)\}$ with permutation. This may have higher computational cost than the method above due to permutation, but is computationally more efficient than estimating p-values directly via permutation or bootstrap, especially when evaluating small p-values.

B.7 Supplementary Simulations for Chapter III

B.7.1 Simulations on One-Sample Covariance Testing

In the following Sections B.7.1.1–B.7.1.5, we present the results of the five simulation settings introduced in Section 3.2.4.

B.7.1.1 Study 1: Empirical Size

In this study, we verify the theoretical results under H_0 in Section 3.2 and the show validity of the adaptive testing procedure across different n and p values. In particular, we fix $n = 100$ and take $p \in \{50, 100, 200, 400, 600, 800, 1000\}$. Then we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \dots, n$, and each $\mathbf{x}_i$ has i.i.d. entries of $\mathcal{N}(0, 1)$ and Gamma$(2, 0.5)$ respectively. The results are summarized in the following Tables B.1 and B.2 respectively.

Table B.1: Empirical type I errors under the Gaussian distribution when $n = 100$.

p	50	100	200	400	600	800	1000
$\mathcal{U}(1)$	0.054	0.055	0.045	0.053	0.048	0.052	0.036
$\mathcal{U}(2)$	0.058	0.058	0.066	0.050	0.071	0.048	0.063
$\mathcal{U}(3)$	0.057	0.066	0.061	0.055	0.051	0.063	0.052
$\mathcal{U}(4)$	0.054	0.067	0.052	0.080	0.053	0.041	0.056
$\mathcal{U}(5)$	0.049	0.054	0.059	0.070	0.045	0.049	0.053
$\mathcal{U}(6)$	0.039	0.057	0.063	0.061	0.056	0.057	0.074
$\mathcal{U}(\infty)$ 1	0.046	0.055	0.049	0.067	0.064	0.042	0.044
$\mathcal{U}(\infty)$ 2	0.040	0.047	0.045	0.056	0.048	0.050	0.048
adpUmin 1	0.056	0.066	0.067	0.064	0.067	0.056	0.051
adpUf 1	0.065	0.083	0.069	0.079	0.063	0.058	0.060
adpUmin 2	0.054	0.069	0.065	0.060	0.062	0.055	0.057
adpUf 2	0.069	0.082	0.065	0.065	0.058	0.057	0.062
Identity	0.055	0.053	0.058	0.053	0.061	0.049	0.053
Sphericity	0.053	0.050	0.058	0.053	0.062	0.049	0.054
LW	0.058	0.051	0.053	0.045	0.067	0.048	0.058
Schott	0.052	0.055	0.050	0.052	0.050	0.044	0.051

In Tables B.1 and B.2, we provide the simulation results of all the single U-statistics with orders in $\{1, \dots, 6\}$. For $\mathcal{U}(\infty)$, we first use the test statistic (3.9) same as in Jiang (2004), which is denoted as "$\mathcal{U}(\infty)$ 1" below. Since the convergence in Jiang (2004) is slow, we use permutation to approximate the distribution in the simulations. We also use the standardized version $M_n^\dagger$ introduced in Remark B.2, which is denoted as "$\mathcal{U}(\infty)$ 2" below. Given "$\mathcal{U}(\infty)$ 1" and "$\mathcal{U}(\infty)$ 2", we apply the adaptive testing with minimum combination and Fisher's method respectively. The results are denoted as "adpUmin1", "adpUf1", "adpUmin2" and "adpUf2" respec-

Table B.2: Empirical type I errors under the Gamma distribution when $n = 100$.

p	50	100	200	400	600	800	1000
$\mathcal{U}(1)$	0.043	0.049	0.054	0.048	0.050	0.049	0.043
$\mathcal{U}(2)$	0.057	0.075	0.062	0.054	0.057	0.055	0.061
$\mathcal{U}(3)$	0.054	0.064	0.050	0.041	0.057	0.051	0.056
$\mathcal{U}(4)$	0.047	0.056	0.061	0.056	0.052	0.053	0.045
$\mathcal{U}(5)$	0.043	0.043	0.054	0.052	0.050	0.053	0.049
$\mathcal{U}(6)$	0.032	0.035	0.059	0.045	0.046	0.053	0.044
$\mathcal{U}(\infty)$ 1	0.052	0.045	0.048	0.053	0.045	0.049	0.055
$\mathcal{U}(\infty)$ 2	0.044	0.052	0.052	0.053	0.044	0.051	0.045
adpUmin 1	0.051	0.054	0.069	0.062	0.049	0.058	0.065
adpUf 1	0.055	0.060	0.075	0.067	0.054	0.058	0.067
adpUmin 2	0.049	0.055	0.068	0.063	0.049	0.059	0.066
adpUf 2	0.063	0.067	0.070	0.058	0.047	0.057	0.061
Identity	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Sphericity	0.088	0.065	0.071	0.056	0.060	0.059	0.050
LW	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Schott	0.051	0.063	0.053	0.053	0.055	0.046	0.060

tively below. In addition, we also compare several methods in the literature. The identity and sphericity tests in Chen et al. (2010) are denoted as "Equal" and "Spher" below; the methods in Ledoit and Wolf (2002) and Schott (2007), which are referred to as "LW" and "Schott" respectively.

Remark B.2. *Besides M_n^* in (3.9), an alternative way to construct $\mathcal{U}(\infty)$ in the framework is to standardize $\hat{\sigma}_{j_1,j_2}$ by its variance $\widehat{\mathrm{var}}(\hat{\sigma}_{j_1,j_2})$. Specifically, following Cai et al. (2013), we take $\widehat{\mathrm{var}}(\hat{\sigma}_{j_1,j_2}) = n^{-1}\sum_{i=1}^{n}\{(x_{i,j_1} - \bar{x}_{j_1})(x_{i,j_2} - \bar{x}_{j_2}) - \hat{\sigma}_{j_1,j_2}\}^2$. Define $M_n^\dagger = \max_{1\le j_1\ne j_2\le p}|\hat{\sigma}_{j_1,j_2}|/\{\widehat{\mathrm{var}}(\hat{\sigma}_{j_1,j_2})\}^{1/2}$ and we take $\mathcal{U}(\infty) = M_n^\dagger$. Theoretically, it can be shown that Theorem 3.2.3 still holds with $\mathcal{U}(\infty) = M_n^\dagger$ (He et al., 2021d, Section B.11). Numerically, we provide the simulations in Section B.7.1, which shows that M_n^* in (3.9) generally has higher power than $M_n^\dagger$.*

B.7.1.2 Study 2

In this section, we provide the simulation results for the second setting in Section 3.2.4. In particular, we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \ldots, n$, and

$\mathbf{x}_i$ follows multivariate Gaussian distribution with mean zero and covariance $\mathbf{\Sigma}_A = (1-\rho)I_p + \rho \mathbf{1}_{p,k_0}\mathbf{1}_{p,k_0}^\intercal$.

Similarly to Figure III.2, we conduct simulations on the adaptive procedure with U-statistics of orders in $\{1, \ldots, 6, \infty\}$. We provide the simulation results of all the single U-statistics and the adaptive procedure, and also compare with some other methods in the literature. We take $(n, p) \in \{(100, 300), (100, 600), (100, 1000)\}$, and provide the results in the following Figures B.1–B.3 respectively.

In Figure B.1, the first 7 plots are simulated with $k_0 \in \{2, 5, 7, 10, 13, 20, 50\}$. Particularly, we include results of $\mathcal{U}(a)$ for $a \in \{1, \ldots, 6, \infty\}$; the adaptive procedure "adpU" by minimum combination of these single U-statistics; identity and sphericity tests in Chen et al. (2010), which are denoted as 'Equal" and "Shper", respectively. We can see that when $k_0 \in \{7, 10, 13\}$, the results of "adpU" are better than all the other test statistics. For other cases, the results of "adpU" are close to the best results of single U-statistics. In addition, we also examine the case when the nonzero off-diagonal elements of $\mathbf{\Sigma}_A$, i.e., σ_{j_1,j_2} with $1 \leq j_1 \neq j_2 \leq k_0$, have same absolute value $|\rho|$, but can be positive or negative with equal probability. The results of powers versus different $|\rho|$ values are given by 8th plot in Figure B.1, which is consistent with Remark III.3 in Section 3.2.2.

In Figures B.2 and B.3, the meanings of the legends are the same as in Tables B.1 and B.2, and are already explained in Section B.7.1.1. We can find similar patterns to that in Figure B.1.

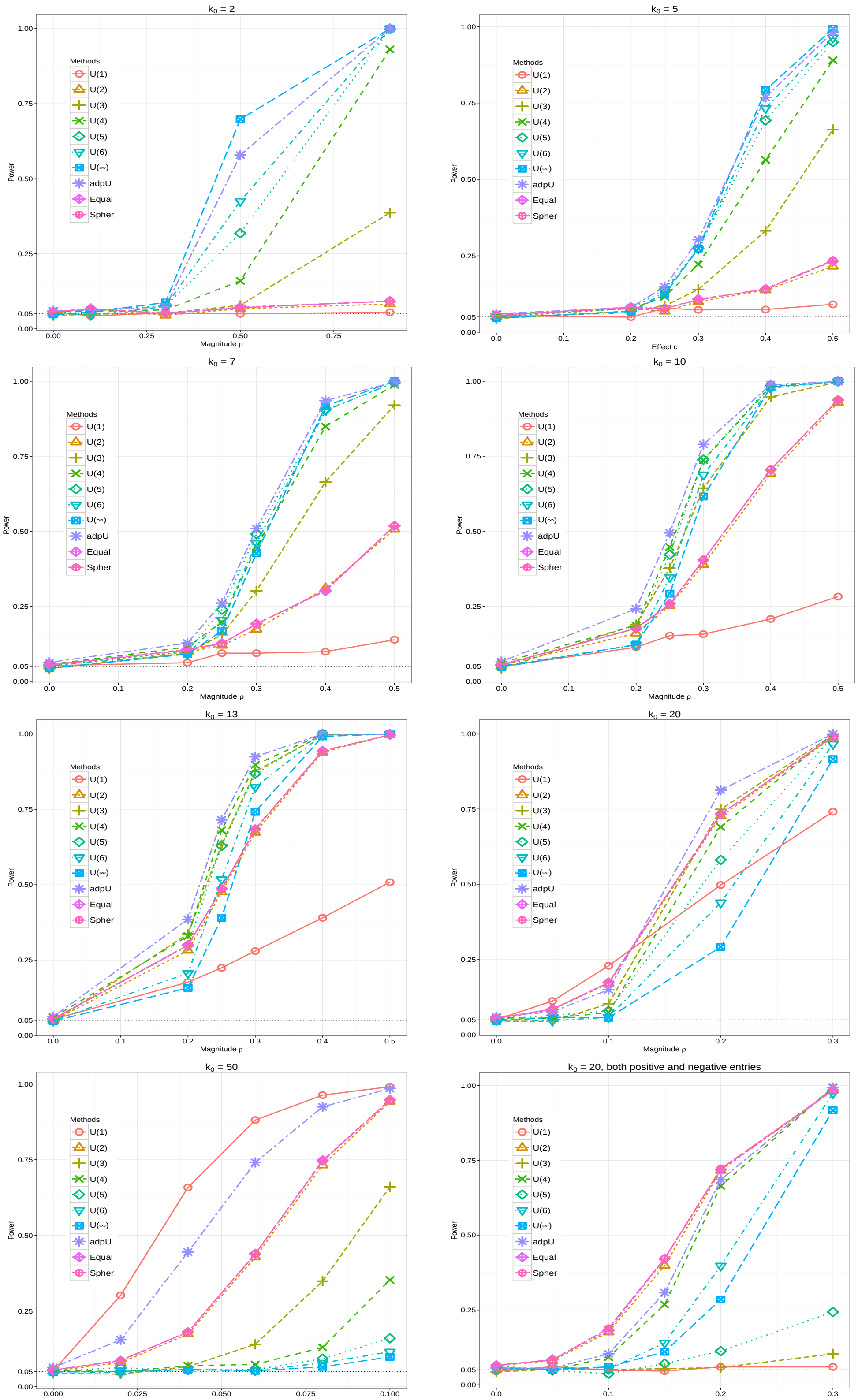


Figure B.1: Study 2 in Section B.7.1.2: Empirical power versus the signal magnitude ρ when $n = 100$, $p = 300$, and $k_0 \in \{2, 5, 7, 10, 13, 20, 50\}$.

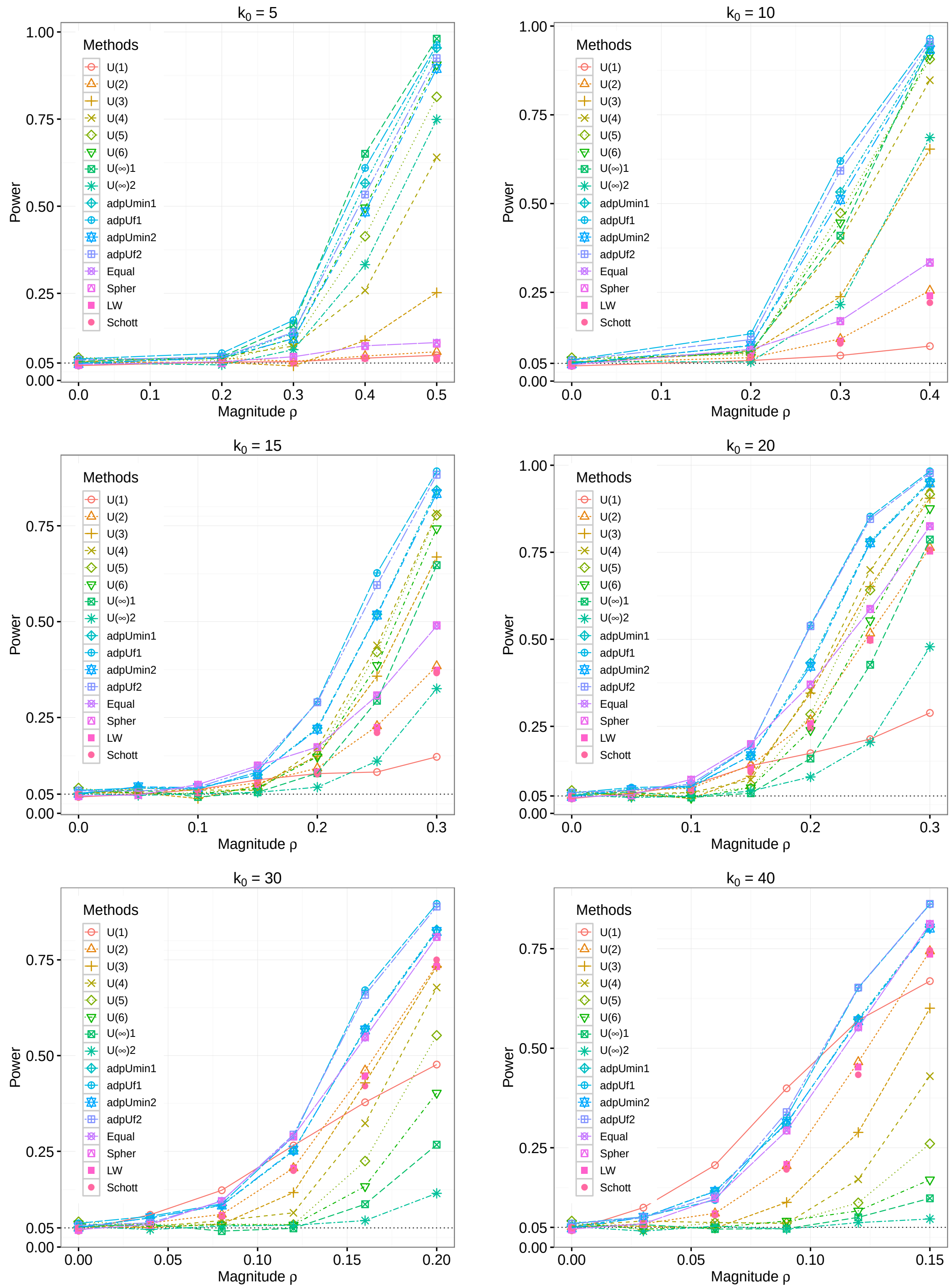


Figure B.2: Study 2 in Section B.7.1.2: Empirical power versus the signal magnitude ρ when $n = 100$, $p = 600$, and $k_0 \in \{5, 10, 15, 20, 30, 40\}$.

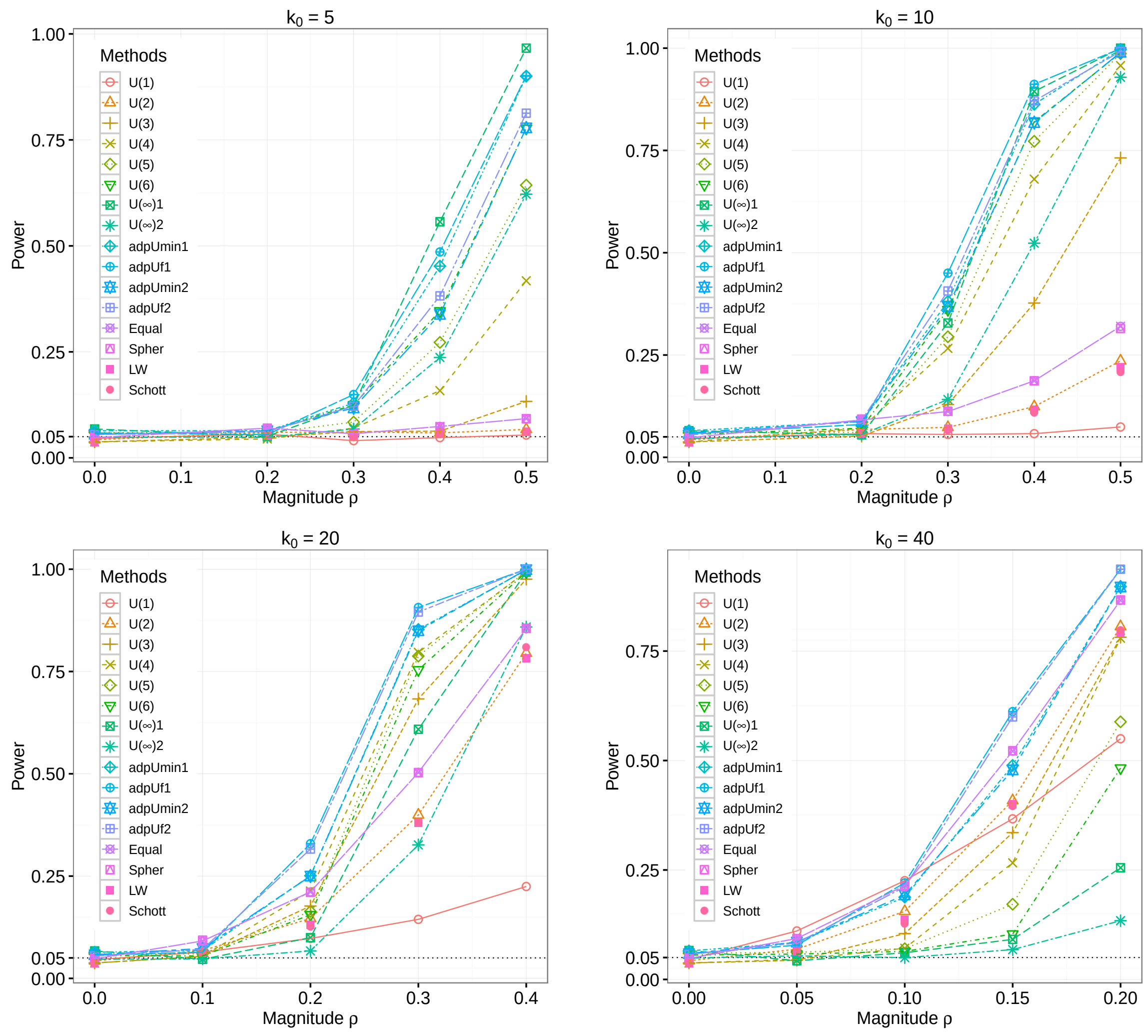


Figure B.3: Study 2 in Section B.7.1.2: Empirical power versus the signal magnitude ρ when $n = 100$, $p = 1000$, and $k_0 \in \{5, 10, 20, 40\}$.

B.7.1.3 Study 3

We provide supplementary simulations for the third setting in Section 3.2.4. In particular, we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \ldots, n$, and $\mathbf{x}_i$ follows multivariate Gaussian distribution with mean zero and covariance $\mathbf{\Sigma}_A$. In this case, $\mathbf{\Sigma}_A$ is symmetric and positive definite, and has the diagonal being all one and only $|J_A|$ random positions being nonzero with value ρ. Note that here ρ represents the magnitude of the alternative signal; and $|J_A|$ represents its sparsity level with a larger value indicating a denser alternative, and vice versa. We let $|J_A|$ and ρ vary to examine how the power changes correspondingly. We take $(n, p) \in \{(100, 600), (100, 1000)\}$, and provide the results in the following Figures B.4–B.5 respectively. The meanings of the legends are the same as in Tables B.1 and B.2, and are already explained in Section B.7.1.1. We observe similar patterns to that in the figures in Section B.7.1.2.

B.7.1.4 Study 4

In this section, we provide the simulation results of the fourth setting in Section 3.2.4. In particular, we generate n i.i.d. p-dimensional $\mathbf{x}_i$ for $i = 1, \ldots, n$, and $\mathbf{x}_i$ follows multivariate Gaussian distribution with mean zero and covariance $\mathbf{\Sigma}_A$. Under this setting, $\mathbf{\Sigma}_A$ is symmetric and positive definite and has the diagonal being all one and $|J_A|$ random positions taking values uniformly in the range $(0, 2\rho)$. Therefore, the nonzero off-diagonal elements in $\mathbf{\Sigma}_A$ are different. Figure B.6 below presents the power versus ρ when $n = 100$ and $p = 1000$. The meanings of the legends are the same as in Tables B.1 and B.2, and are already explained in Section B.7.1.1. We observe similar patterns to that in the figures in Section B.7.1.2.

B.7.1.5 Study 5

In this section, we compare our methods with the methods in Chen et al. (2010) following their multivariate models. Specifically, for each $i = 1, \ldots, n$, $\mathbf{x}_i = \Xi \mathbf{z}_i + \boldsymbol{\mu}$,

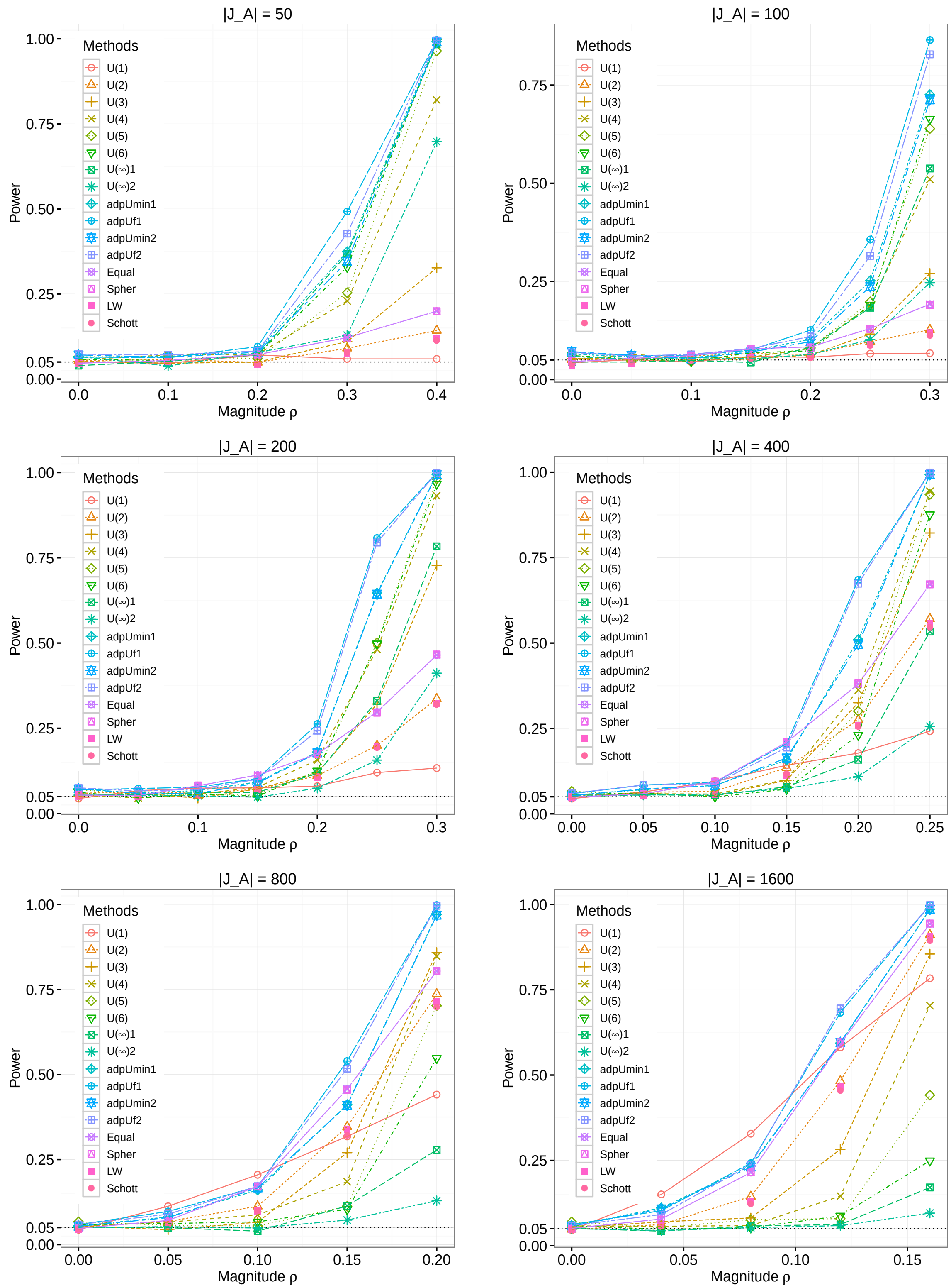


Figure B.4: Study 3 in Section B.7.1.3: Empirical power versus the signal magnitude ρ when $n = 100, p = 600$, and $|J_A| \in \{50, 100, 200, 400, 800, 1600\}$.

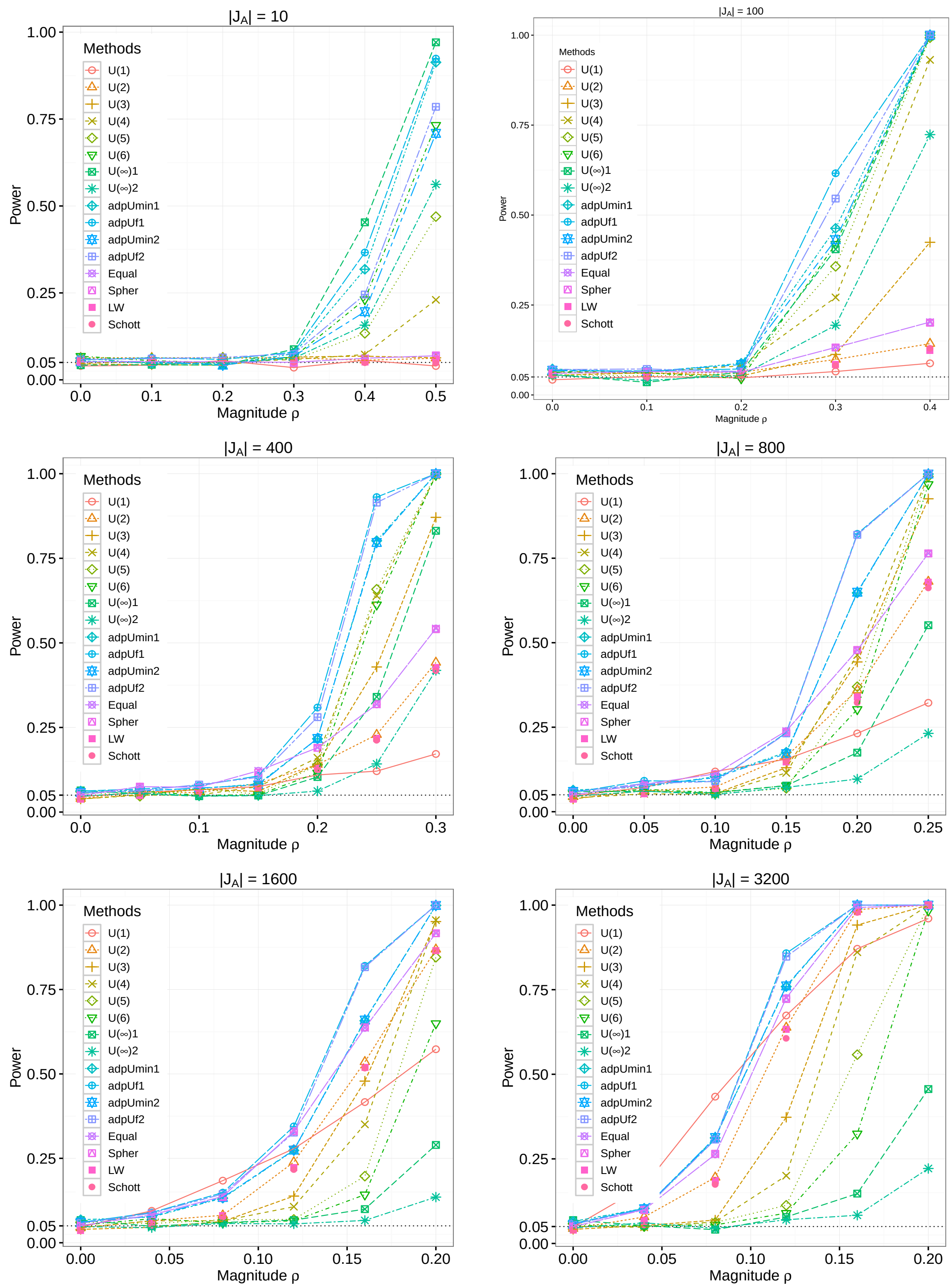


Figure B.5: Study 3 in Section B.7.1.3: Empirical power versus the signal magnitude ρ when $n = 100, p = 1000$, and $|J_A| \in \{10, 100, 400, 800, 1600, 3200\}$.

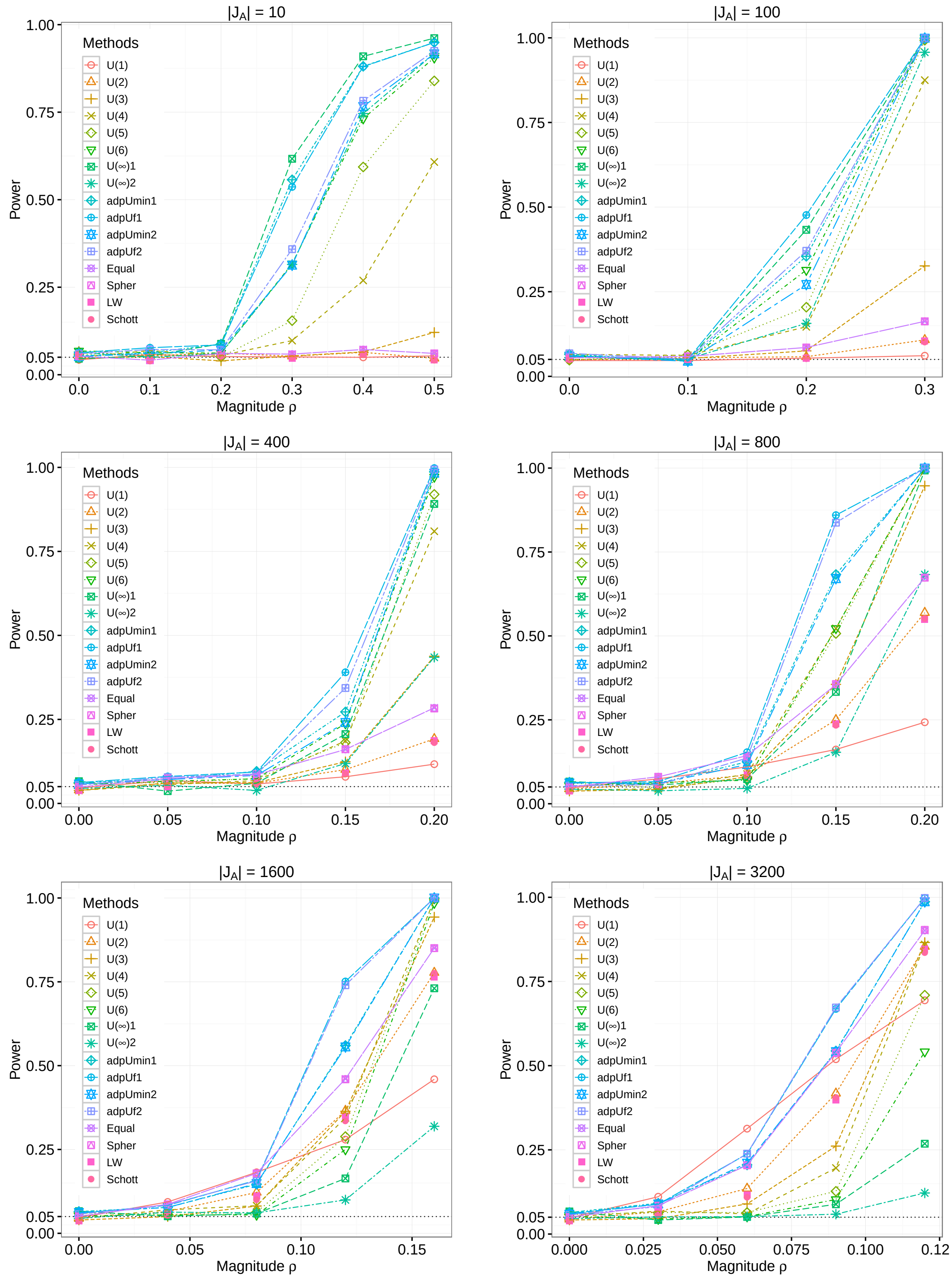


Figure B.6: Study 4 in Section B.7.1.4: Empirical power versus the signal magnitude ρ when $n = 100, p = 1000$, and $|J_A| \in \{10, 100, 400, 800, 1600, 3200\}$.

where Ξ is a matrix of dimension $p \times m$ with $m \geq p$. Under null hypothesis, $m = p$, $\Xi = I_p$ $\boldsymbol{\mu} = \mu_0 \mathbf{1}_p$ with $\mu_0 = 2$; under alternative hypothesis, $m = p + 1$, $\boldsymbol{\mu} = 2(\sqrt{1-\rho} + \sqrt{2\rho})\mathbf{1}_p$, $\Xi = (\sqrt{1-\rho} I_p, \sqrt{2\rho}\mathbf{1}_p)$, thus $\boldsymbol{\Sigma} = (1-\rho)I_p + 2\rho \mathbf{1}_p \mathbf{1}_p^\intercal$. Two settings are examined: (1) $\mathbf{z}_i$'s are i.i.d. multivariate Gaussian random vectors with mean $\mathbf{0}$ and covariance I_p; (2) $\mathbf{z}_i = (z_{i,1}, \ldots, z_{i,m})^\intercal$ consists of i.i.d. random variables $z_{i,j}$ which are standardized Gamma$(4, 0.5)$ random variables so that $\mathbf{z}_i$ has mean $\mathbf{0}$ and covariance I_p.

To mimic "large p, small n" situation, Chen et al. (2010) sets dimension $p = c_1 \exp(n^\eta) + c_2$, where $\eta = 0.4$, for $(c_1, c_2) = (1, 10)$ and $(c_1, c_2) = (2, 0)$ respectively. In particular, we consider $(n, p) \in \{(40, 159), (40, 331), (80, 159), (80, 331), (80, 642)\}$. The results are based on 1000 simulations and the nominal significance level of the tests is 5%.

In the Tables B.3–B.10, results outside and inside parentheses are calculated from parametric-permutation- and asymptotics-based methods, respectively. To be specific, psarametric-permutation-based method means estimating p-values or powers by permutation; and asymptotic-based method uses the asymptotic theoretical results and is described in Section 3.2.3. For each $a \in \{1, \ldots, 6, \infty\}$, the row of "$\mathcal{U}(a)$" has results using the single test statistic $\mathcal{U}(a)$; and the row of "adpU" is obtained by the adaptive testing procedure which combines all single candidate U-statistics in the tables using the minimum combination. In addition, "Ident" and "Spher" rows denote the identity and sphericity tests in Chen et al. (2010) separately.

In the tables B.3–B.8, we find that the empirical sizes of most tests are close to the nominal level, except $\mathcal{U}(\infty)$ due to the slow convergence to extreme value distribution as pointed out in Hall (1979). "Ident" and "Spher" tests perform similarly to $\mathcal{U}(2)$ in both settings. This is reasonable because they are all sum-of-squares-type statistics. Moreover, for the ρ's examined, $\mathcal{U}(1)$ has higher power than $\mathcal{U}(2)$, as the constructed alternative is very dense and only has positive entries. In addition, "adpU" achieves high power for different cases, and its power converges to 1, as one of the test statistics has power converging to 1. In Tables B.9 and B.10, data are standardized with sample mean and variance. It can be seen that methods in Chen et al. (2010) perform poorly in this case. Other than this, the results follow similar patterns to results in other tables.

Table B.3: Empirical type I errors and power (%) under the simulation setting (1) of Study 5 in Section B.7.1.5 when $n = 80$ and $p = 331$.

ρ	0	0.001	0.002	0.003	0.004
$\mathcal{U}(1)$	4.4 (4)	93.4 (90.6)	100 (99.9)	100 (100)	100 (100)
$\mathcal{U}(2)$	5 (5.6)	5.5 (6)	7.2 (5.9)	13.1 (10.2)	19.7 (14.4)
$\mathcal{U}(3)$	5.4 (6.1)	4.5 (4)	6.3 (5.4)	6.9 (4.5)	9 (5.4)
$\mathcal{U}(4)$	4.7 (5.1)	6 (5.4)	3.7 (4.6)	4.2 (5.3)	6 (4.8)
$\mathcal{U}(5)$	5.4 (6.3)	4.9 (4.7)	5.3 (5.6)	6 (5.7)	6.1 (5.1)
$\mathcal{U}(6)$	4.6 (4.9)	5.8 (5.4)	4.9 (4.5)	5.2 (4.8)	4.8 (5)
$\mathcal{U}(\infty)$	4.7 (0.3)	5 (0.6)	5.5 (0.7)	5.1 (0.4)	5.9 (0.8)
aSPU	5 (5.4)	81 (81.8)	99.4 (99.4)	100 (100)	100 (100)
Ident	5.5	5.7	8.2	14.4	21.8
Spher	5.6	5.7	8.1	14.2	21.4

Table B.4: Empirical type I errors and power (%) under the simulation setting (2) of Study 5 in Section B.7.1.5 when $n = 80$ and $p = 331$.

ρ	0	0.001	0.002	0.003	0.004
$\mathcal{U}(1)$	5.3 (4.6)	56.7 (50.3)	92.5 (89.3)	99.3 (99.1)	100 (99.8)
$\mathcal{U}(2)$	5.4 (5)	5.5 (5.7)	6.9 (5.4)	7.7 (5.8)	11.4 (7.3)
$\mathcal{U}(3)$	5.6 (5.4)	4.5 (3.5)	5.7 (4)	5.8 (4.8)	7.2 (5.1)
$\mathcal{U}(4)$	4.8 (3.9)	4.9 (4.1)	4.9 (5)	6.5 (6.8)	4.9 (5.1)
$\mathcal{U}(5)$	6.1 (5.1)	5.6 (6.1)	5.1 (5.2)	5.5 (5.7)	5.2 (5.5)
$\mathcal{U}(6)$	6.4 (5.6)	5.4 (4.1)	5.1 (5.3)	5.1 (5.4)	5.8 (5.3)
$\mathcal{U}(\infty)$	5.5 (3)	5.3 (2.5)	6 (2.8)	5.5 (2.8)	6.8 (3.1)
adpU	6.4 (6.5)	35 (36.3)	78.7 (79.2)	96.1 (96.1)	99.5 (99.6)
Ident	6.7	6.5	7.4	9.2	13.5
Spher	6.2	6.2	7	9.1	12.9

Table B.5: Empirical Type I errors and power (%) under the simulation setting (1) of Study 5 in Section B.7.1.5 when $n = 40$ and $p = 159$.

ρ	0	0.0005	0.001	0.0015	0.002	0.0025
$\mathcal{U}(1)$	5.8 (4.6)	16.6 (13.6)	36.5 (32.3)	57.4 (51.3)	69.2 (65.1)	83.3 (80)
$\mathcal{U}(2)$	5.2 (4.9)	4.6 (3.1)	4.6 (5.6)	5.3 (4.5)	5.5 (4.8)	5.9 (4.8)
$\mathcal{U}(3)$	4.9 (4.8)	5.8 (5.4)	5.6 (5.6)	5.6 (4.9)	4.6 (4.7)	5.6 (5)
$\mathcal{U}(4)$	4.6 (5.7)	4.2 (4.1)	5.6 (4.6)	4.7 (4.6)	4.5 (5.1)	5.3 (4.9)
$\mathcal{U}(5)$	5.5 (5.6)	5.3 (6.2)	5.7 (4.9)	3.1 (3.1)	4.7 (4.4)	5.5 (5.4)
$\mathcal{U}(6)$	4.4 (4.3)	4.8 (4.6)	4.4 (4.7)	4.3 (4.3)	4.8 (4.6)	5 (4.2)
$\mathcal{U}(\infty)$	5.1 (0.1)	5.1 (0.1)	4.2 (0)	4.6 (0.1)	4.6 (0)	5.5 (0.1)
adpU	5.7 (5.8)	8.9 (10.6)	18.5 (21.1)	31.5 (34.2)	47.4 (50.8)	63.2 (66.2)
Ident	5.8	5.3	5.9	6.8	6.8	7.1
Spher	5.8	5.1	5.7	6.5	6.5	7.2

Table B.6: Empirical Type I errors and power (%) under the simulation setting (1) of Study 5 in Section B.7.1.5 when $n = 40$ and $p = 331$.

ρ	0	0.0025	0.005	0.01	0.015	0.02
$\mathcal{U}(1)$	5.9 (5.4)	99.4 (99.3)	100 (100)	100 (100)	100 (100)	100 (100)
$\mathcal{U}(2)$	5.1 (4.4)	7 (6.3)	15.5 (10.7)	65.8 (60)	95.1 (93.1)	99.3 (98.7)
$\mathcal{U}(3)$	5.4 (5.5)	7.6 (4.6)	13 (7.5)	26.3 (19.7)	53.9 (44.1)	76.9 (68.9)
$\mathcal{U}(4)$	4.8 (5.1)	4.9 (5.4)	6.8 (5.6)	6.3 (6.6)	11.4 (7.7)	14.4 (11.7)
$\mathcal{U}(5)$	5.9 (4.8)	5.5 (4.9)	7 (6.6)	5.6 (4.9)	8.6 (7.3)	8.5 (8.2)
$\mathcal{U}(6)$	4.1 (4.9)	3.4 (4.5)	6.8 (4.6)	4.8 (6.5)	5.5 (6.6)	8 (8.6)
$\mathcal{U}(\infty)$	4.2 (0)	4.1 (0)	6.1 (0)	4.9 (0)	6.6 (0)	7.3 (0.1)
adpU	5.2 (5.8)	97.5 (98.5)	100 (100)	100 (100)	100 (100)	100 (100)
Ident	6.2	8.3	19.2	68	95.5	99.3
Spher	6.3	8.2	18.6	67.6	95.4	99.3

Table B.7: Empirical type I errors and power (%) under the simulation setting (1) of Study 5 in Section B.7.1.5 when $n = 80$ and $p = 159$.

ρ	0	0.0025	0.005	0.01	0.015	0.02
$\mathcal{U}(1)$	5.7 (4.7)	98.1 (97)	100 (100)	100 (100)	100 (100)	100 (100)
$\mathcal{U}(2)$	6.2 (5.1)	6.8 (5.5)	16.5 (11.4)	68.4 (60.6)	96.7 (94.7)	100 (99.9)
$\mathcal{U}(3)$	6 (4.7)	6.2 (5.5)	7.4 (5.9)	15.2 (9.2)	34.8 (26.2)	69.2 (61.4)
$\mathcal{U}(4)$	5.4 (5.6)	4 (3.8)	4.7 (4.2)	7.6 (7.1)	10.6 (9)	18.2 (15.7)
$\mathcal{U}(5)$	4.5 (4.9)	4.6 (4.2)	4.8 (4.5)	5.3 (5.3)	9.6 (7.6)	13.1 (13)
$\mathcal{U}(6)$	5.6 (5.3)	3.9 (4.7)	4 (3.3)	5.3 (4.9)	8.7 (8)	12 (12.4)
$\mathcal{U}(\infty)$	4.5 (0.8)	6.1 (1.1)	4.9 (1.4)	5.4 (1.7)	8 (1.5)	10.7 (3.3)
adpU	5.7 (7)	91.8 (92.6)	99.8 (99.8)	100 (100)	100 (100)	100 (100)
Ident	6.7	7.8	18.5	71.1	97.3	100
Spher	6.7	7.2	18	69.6	97	100

Table B.8: Empirical type I errors and power (%) under the simulation setting (1) of Study 5 in Section B.7.1.5 when $n = 80$ and $p = 642$.

ρ	0	0.0025	0.005	0.01	0.015	0.02
$\mathcal{U}(1)$	5.8 (4.8)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
$\mathcal{U}(2)$	6.4 (6.2)	17.9 (12.7)	71.2 (63.4)	99.8 (99.8)	100 (100)	100 (100)
$\mathcal{U}(3)$	5.2 (5.6)	6.2 (3.6)	19.3 (13.3)	68.4 (57.3)	96.4 (94)	99.8 (99.6)
$\mathcal{U}(4)$	5.2 (5.2)	6.2 (6.4)	5.2 (5.2)	8.5 (6.4)	25 (18.3)	57.9 (51.7)
$\mathcal{U}(5)$	6.4 (4.6)	5 (5.2)	6.4 (5.4)	7.8 (7.2)	11.7 (9.9)	21.1 (16.9)
$\mathcal{U}(6)$	4 (4.2)	5.8 (6.4)	6 (6)	4.2 (5.2)	9.3 (10.3)	13.1 (15.3)
$\mathcal{U}(\infty)$	4.4 (0.6)	5 (0.2)	5.6 (0.4)	7 (0.8)	9.3 (0.8)	15.3 (0.6)
adpU	6 (4.2)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)
Ident	6.8	18.9	72.6	100	100	100
Spher	6.6	18.7	72.6	100	100	100

Table B.9: Empirical type I errors and power (%) under the simulation setting (2) of Study 5 in Section B.7.1.5 when $n = 80$ and $p = 159$.

ρ	0	0.0005	0.001	0.002	0.003	0.004
$\mathcal{U}(1)$	4.9 (4.2)	26.1 (20.4)	57.1 (49.7)	95.2 (93.1)	99.9 (99.8)	100 (99.9)
$\mathcal{U}(2)$	4.9 (4.4)	3.9 (5.3)	5.9 (5.2)	6.7 (4.8)	8.3 (5.6)	12.2 (7.7)
$\mathcal{U}(3)$	5.4 (5.2)	4.7 (5.3)	4.3 (4.1)	6 (4)	5.9 (5.1)	7 (5)
$\mathcal{U}(4)$	5.4 (4.9)	5.5 (5.2)	4.8 (4.8)	5.9 (6.3)	6.7 (7.2)	4.6 (4.6)
$\mathcal{U}(5)$	7.3 (6.2)	5.4 (5.6)	5.8 (6.5)	5.3 (6.3)	5.8 (5.5)	5.6 (5.6)
$\mathcal{U}(6)$	6.5 (5.6)	4.9 (5)	5.5 (5.3)	4.9 (5.2)	5.5 (5.4)	4.2 (4.7)
$\mathcal{U}(\infty)$	5.9 (3)	5.7 (2.1)	5.8 (2.5)	5.7 (2.6)	5.5 (2.9)	6.7 (3.3)
adpU	5.7 (5)	12.1 (13.1)	34.8 (34.6)	81.9 (82.6)	98.1 (98.1)	99.9 (99.8)
Ident	0.2	0.1	0.1	0.2	0.1	0.1
Spher	0.2	0.1	0.1	0.2	0	0.1

Table B.10: Empirical type I errors and power (%) under the simulation setting (2) of Study 5 in Section B.7.1.5 when $n = 80$ and $p = 642$.

ρ	0	0.0005	0.001	0.002	0.003	0.004
$\mathcal{U}(1)$	2.8 (2.2)	94.2 (93)	100 (100)	100 (100)	100 (100)	100 (100)
$\mathcal{U}(2)$	5.8 (4.2)	4.2 (4.8)	6 (5.6)	11.9 (7.2)	22.3 (14.5)	45.9 (36.2)
$\mathcal{U}(3)$	3.6 (3.8)	5.4 (5.2)	7.2 (5)	6 (3.6)	11.9 (7.6)	15.1 (9.3)
$\mathcal{U}(4)$	4.4 (4.4)	4.6 (4.4)	6.4 (6.2)	4.8 (3.8)	5.4 (5.2)	7 (6.2)
$\mathcal{U}(5)$	7 (5.6)	6 (5)	6.2 (5.4)	7 (6.2)	6.6 (5.4)	7.4 (5.6)
$\mathcal{U}(6)$	7 (5.4)	5 (4.6)	4.6 (5.6)	6.8 (7.2)	5.4 (4.6)	5.6 (5.8)
$\mathcal{U}(\infty)$	4.8 (2.2)	6.2 (2.4)	4.8 (0.8)	6.2 (3)	6.4 (2.6)	5.2 (1.6)
adpU	5 (4)	84.5 (85.9)	100 (100)	100 (100)	100 (100)	100 (100)
Ident	0	0.4	0.2	0.4	2.4	8.3
Spher	0	0.4	0.2	0.4	2.4	7.8

B.7.2 Simulation on Testing Coefficients in the GLM

In this study, we conduct simulations for generalized linear model considering the following model

$$y_i = \mathbf{z}_i^\intercal \boldsymbol{\alpha} + \mathbf{x}_i^\intercal \boldsymbol{\beta} + \epsilon_i, \tag{B.213}$$

for $i = 1, \ldots, n$. We generate i.i.d. $\mathbf{x}_i$ from the multivariate normal distribution $\mathcal{N}(0, \Sigma)$. We show the results with an equal variance and a first-order autoregressive correlation matrix case, that is, $\Sigma = (0.4^{|i-j|})$. We further generate $\mathbf{z}_i$ of two covariates with entries i.i.d. from standard normal distribution $\mathcal{N}(0, 1)$, and ϵ_i are the random errors following i.i.d. normal distribution $\mathcal{N}(0, 0.5)$. In (B.213), we take $\boldsymbol{\alpha} = (0.3, 0.3)^\intercal$, $\boldsymbol{\beta} = \mathbf{0}$ or $\neq \mathbf{0}$ corresponded to the null hypothesis H_0 and the alternative hypothesis H_A, respectively. Under H_A, $\lfloor ps \rfloor$ elements in $\boldsymbol{\beta}$ are set to be non-zero, where $s \in [0, 1]$ controls signal sparsity. We vary s to mimic varying sparsity situations, from sparse to dense signals with $s \in \{0.001, 0.1, 0.3, 0.7, 0.9\}$. The positions of non-zero elements in $\boldsymbol{\beta}$ are assumed to be uniformly distributed in $\{1, 2, \ldots, p\}$, and their values are constant c, where c is the effect of signals that vary in the simulations. The results are based on 1000 simulations with 5% nominal significance level, $n = 500$ and $p = 1000$. We summarized the results in Figure B.7. It shows similar patterns as in Study I.

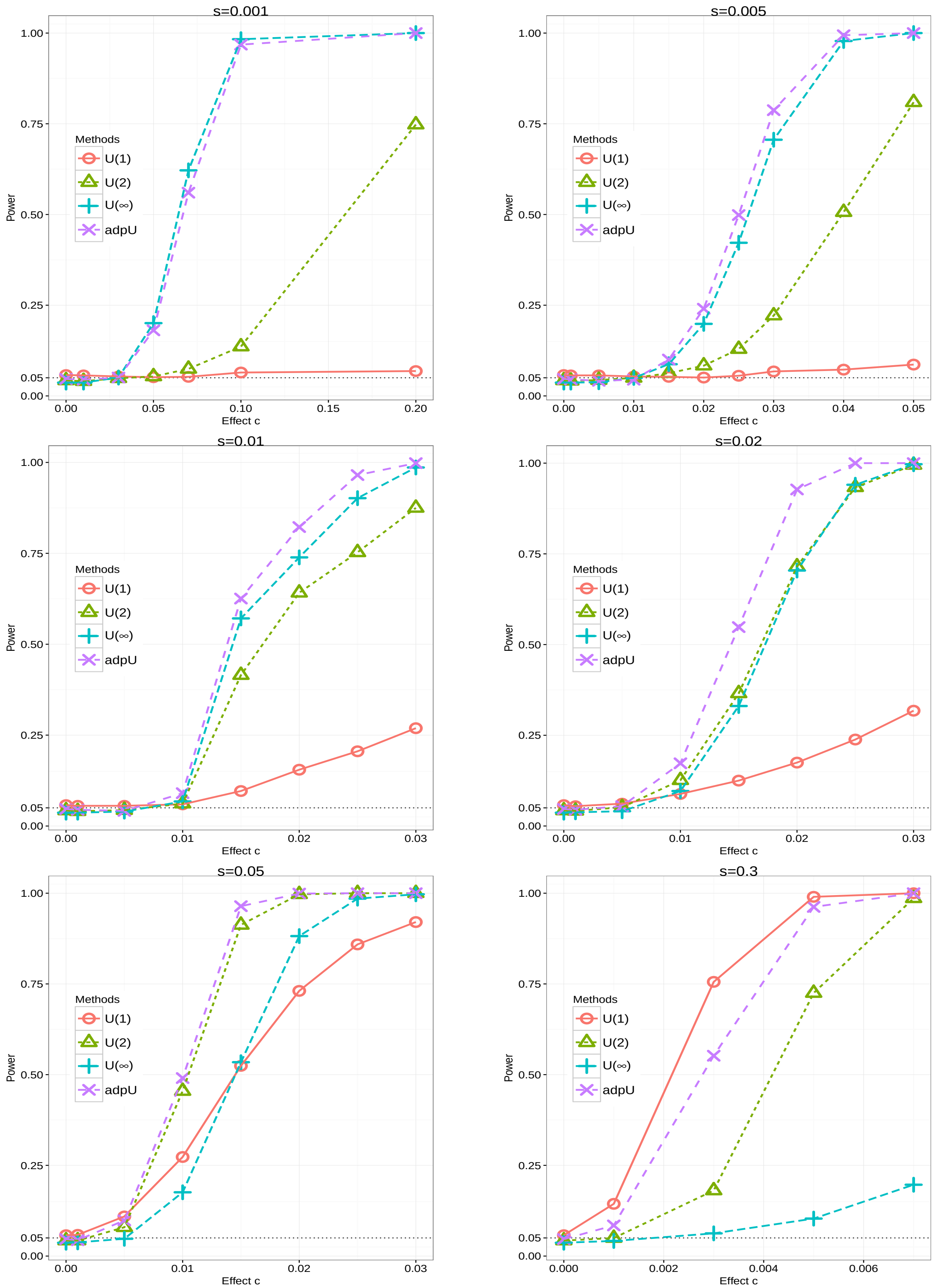


Figure B.7: Power comparison of different tests for the coefficients of the generalized linear model (B.213).

B.7.3 Simulations on Two-sample Covariance Tests

B.7.3.1 Empirical Size Under H_0

In this section, we examine the empirical Type I errors of the proposed the adaptive testing procedure and compare it with the other methods.

We follow the simulation settings in Yang and Pan (2017). In particular, let $A(s)$ be the $s \times s$ covariance matrix of MA(1) model with the parameter $\theta_1 = 0.4$. In addition, $B = 0.7I_{p-s}$ is a $(p-s) \times (p-s)$ scaled identity matrix. We then define the matrix $Q(s) = \mathrm{BlkDiag}(A(s), B)$, where "BlkDiag" indicates a block diagonal matrix. We take $s = p^{1/2}$ and $n = 100$, and consider $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y = Q(s)$. The results are presented in Table B.11.

In Table B.11, we provide the simulation results of the single U-statistics $\mathcal{U}(a)$ with $a \in \{1, \ldots, 6\}$. In addition, we provide the simulation results of $\mathcal{U}(\infty)$ using permutation and the asymptotic distribution in Cai et al. (2013), which are denoted as "$\mathcal{U}(\infty)$ permutation" and "$\mathcal{U}(\infty)$ Tony" respectively. Given the results of $\mathcal{U}(1), \ldots, \mathcal{U}(6)$ and "$\mathcal{U}(\infty)$ (permutation)", "adpUmin 1" and "adpUf 1" represent the results of the adaptive testing procedure using minimum combination and Fisher's method respectively. Similarly, given the results of $\mathcal{U}(1), \ldots, \mathcal{U}(6)$ and "$\mathcal{U}(\infty)$ (Tony)", "adpUmin 2" and "adpUf 2" represent the results of the adaptive testing procedure using minimum combination and Fisher's method respectively. Moreover, "Schott", "Sriva" and "Chen" represent the methods in Schott (2007); Srivastava and Yanagihara (2010) and Li and Chen (2012), respectively. In addition, we denote the tests without and with Micro term in Yang and Pan (2017) as "Pan1" and "Pan2" respectively. The tests in Yang and Pan (2017) are time-consuming. Therefore we only provide the simulation results at $p = 50$, which takes about 100 times the time of the proposed adaptive testing procedure.

Based on our simulation results, we find that the empirical Type I errors of the single U-statistics are close the nominal levels, which verifies the theoretical results of Theorem 3.4.1. Moreover, comparing "$\mathcal{U}(\infty)$ (permutation)" and "$\mathcal{U}(\infty)$ (Tony)", we find that using the asymptotic distribution in Cai et al. (2013) gives conservative Type I errors that are smaller than the nominal levels. In addition, by examining the results of minimum combination and Fisher's method, we find that both of the two methods give empirical Type I errors that are close to the nominal level, while the Fisher's method may have slight size inflation compared to the minimum combination.

Table B.11: Empirical Type-I errors of the two-sample covariance tests when $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y = Q(s)$, $n = 100$, and $s = p^{1/2}$.

p	50	100	200	300
$\mathcal{U}(1)$	0.052	0.055	0.040	0.039
$\mathcal{U}(2)$	0.051	0.060	0.053	0.047
$\mathcal{U}(3)$	0.048	0.061	0.054	0.054
$\mathcal{U}(4)$	0.039	0.059	0.067	0.053
$\mathcal{U}(5)$	0.056	0.046	0.041	0.066
$\mathcal{U}(6)$	0.045	0.044	0.041	0.044
$\mathcal{U}(\infty)$ (permutation)	0.047	0.042	0.049	0.052
adpUmin 1	0.043	0.057	0.059	0.053
adpUf 1	0.076	0.081	0.060	0.076
$\mathcal{U}(\infty)$ (Tony)	0.018	0.024	0.016	0.013
adpUmin 2	0.044	0.056	0.059	0.051
adpUf 2	0.051	0.056	0.040	0.050
Chen	0.050	0.049	0.049	0.050
Sriva	0.166	0.002	0.000	0.000
Schott	0.074	0.119	0.236	0.418
Pan1	0.055	NA	NA	NA
Pan2	0.058	NA	NA	NA

B.7.3.2 Power under H_A

In this section, we examine the power of the two-sample covariance tests, where we follow the covariance matrix models in Yang and Pan (2017). In particular, let $H(\tau_0, \tau_1, r) = (h_{i,j})_{p\times p}$, where $h_{i,j} = 0$ except $h_{i,i} = \tau_0$, $i = 1, \ldots, r$ and $h_{i,i+1} = h_{i,i-1} = \tau_1$, $i = 1, \ldots, r-1$. Here τ_0 and τ_1 are used to measure the level of faint alternatives and r is used to measure the sparsity level of alternative. We fix $\mathbf{\Sigma}_x = I_p$, the $p \times p$ identity matrix, and examine the following three representative covariance matrix models of $\mathbf{\Sigma}_y$.

Model 1: (Extreme faint, $\tau_0 = 0.04, \tau_1 = 0.2, r = p$). $\mathbf{\Sigma}_y = I_p + H(0.04, 0.2, p)$. This matrix can also be considered as the covariance matrix of MA(1) model with the parameter $\theta_1 = 0.2$, which is also used in Li and Chen (2012).

Model 2: (Extreme sparse, $\tau_0 = 1, \tau_1 = 1.5, r = 2$). $\mathbf{\Sigma}_y = I_p + H(1, 1.5, 2)$. This model only has four large disturbances compared with $\mathbf{\Sigma}_x$, which is regarded as the extreme sparse (ES) alternative.

Model 3: (Reasonable faint and sparse, $\tau_0 = 0.3, \tau_1 = 0.3, r = p/10$) $\mathbf{\Sigma}_y = I_p + H(0.3, 0.3, p/10)$. The value of r here is between 2 (in Model 2) and p (in Model 1), which is regarded as a moderately sparse setting.

Under each model above, we take $n = 100$, $p \in \{50, 100, 200, 300\}$, and provide the simulation results of the Models 1–3 in the Tables B.12–B.14 respectively. The explanation of each row are the same as in Table B.11, which is given in Section B.7.3.1. Similarly, we note that the tests in Yang and Pan (2017) are very time-consuming. Therefore for "Pan 1" and "Pan 2", we only provide the simulation results at $p = 50$, which takes about 100 times the time of the proposed adaptive testing procedure.

We then analyze the simulation results. Model 1 is the extreme faint case and $\mathbf{\Sigma}_y - \mathbf{\Sigma}_x$ is dense. We find that under this case, the U-statistics of small orders,

Table B.12: Empirical power of the two-sample covariance tests under Model 1 (Extreme faint) when $n = 100$.

p	50	100	200	300
$\mathcal{U}(1)$	0.397	0.389	0.408	0.416
$\mathcal{U}(2)$	0.445	0.458	0.456	0.484
$\mathcal{U}(3)$	0.290	0.309	0.354	0.371
$\mathcal{U}(4)$	0.197	0.211	0.199	0.205
$\mathcal{U}(5)$	0.244	0.397	0.752	0.855
$\mathcal{U}(6)$	0.054	0.052	0.054	0.091
$\mathcal{U}(\infty)$ (permutation)	0.066	0.062	0.044	0.029
adpUmin 1	0.478	0.511	0.692	0.783
adpUf 1	0.600	0.648	0.843	0.886
$\mathcal{U}(\infty)$ (Tony)	0.091	0.072	0.087	0.072
adpUmin 2	0.480	0.513	0.691	0.781
adpUf 2	0.619	0.669	0.855	0.903
Chen	0.573	0.574	0.569	0.623
Sriva	0.513	0.586	0.598	0.569
Schott	0.667	0.731	0.888	0.956
Pan1	0.640	NA	NA	NA
Pan2	0.669	NA	NA	NA

Table B.13: Empirical power of the two-sample covariance tests under Model 2 (Extreme sparse) when $n = 100$.

p	50	100	200	300
$\mathcal{U}(1)$	0.068	0.056	0.048	0.049
$\mathcal{U}(2)$	0.725	0.364	0.122	0.086
$\mathcal{U}(3)$	0.993	0.960	0.850	0.660
$\mathcal{U}(4)$	1.000	0.997	0.988	0.956
$\mathcal{U}(5)$	0.934	0.874	0.803	0.682
$\mathcal{U}(6)$	0.972	0.960	0.935	0.914
$\mathcal{U}(\infty)$ (permutation)	0.966	0.919	0.852	0.772
adpUmin 1	1.000	0.992	0.984	0.959
adpUf 1	1.000	0.996	0.989	0.970
$\mathcal{U}(\infty)$ (Tony)	0.999	1.000	0.997	1.000
adpUmin 2	1.000	0.997	0.993	0.995
adpUf 2	1.000	0.999	0.992	0.992
Chen	0.800	0.457	0.196	0.127
Sriva	0.787	0.433	0.166	0.101
Schott	0.864	0.640	0.550	0.654
Pan1	0.673	NA	NA	NA
Pan2	0.694	NA	NA	NA

Table B.14: Empirical power of the two-sample covariance tests under Model 3 (Reasonable faint and sparse) when $n = 100$.

p	50	100	200	300
$\mathcal{U}(1)$	0.072	0.067	0.069	0.070
$\mathcal{U}(2)$	0.090	0.096	0.096	0.083
$\mathcal{U}(3)$	0.155	0.151	0.152	0.145
$\mathcal{U}(4)$	0.175	0.162	0.162	0.154
$\mathcal{U}(5)$	0.347	0.582	0.868	0.946
$\mathcal{U}(6)$	0.308	0.494	0.732	0.854
$\mathcal{U}(\infty)$ (permutation)	0.028	0.034	0.027	0.018
adpUmin 1	0.337	0.496	0.797	0.901
adpUf 1	0.355	0.535	0.802	0.910
$\mathcal{U}(\infty)$ (asymptotic)	0.254	0.319	0.409	0.403
adpUmin 2	0.348	0.508	0.798	0.901
adpUf 2	0.426	0.620	0.862	0.940
Chen	0.138	0.149	0.153	0.144
Sriva	0.092	0.096	0.097	0.100
Schott	0.189	0.283	0.486	0.712
Pan1	0.167	NA	NA	NA
Pan2	0.186	NA	NA	NA

e.g., $\mathcal{U}(1)$ and $\mathcal{U}(2)$ are powerful. The tests based on the sum-of-squares type statistics including "Chen", "Sriva" and "Schott" are also powerful under this case. Our proposed adaptive testing procedure using Fisher's method has comparable power performance to "Pan 1" and "Pan 2", and is computationally more efficient. Model 2 is the extreme sparse case. Under this case, we find that generally U-statistics of higher orders, e.g., $\mathcal{U}(4)$ and $\mathcal{U}(\infty)$, are more powerful than the U-statistics of smaller orders, e.g., $\mathcal{U}(1)$ and $\mathcal{U}(2)$. Model 3 is the moderately faint and sparse case. Under this case, we can see that a finite-order U-statistic $\mathcal{U}(5)$ is the most powerful one. Neither the maximum-type test statistic $\mathcal{U}(\infty)$ and the sum-of-squares type test statistic $\mathcal{U}(2)$, "Chen", "Sriva" and "Schott" are very powerful. Tests in Yang and Pan (2017) considering only faint or sparse alternatives are not very powerful under this case. On the other hand, the proposed adaptive testing procedure maintains high power under this case.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abramowitz, M., and I. A. Stegun (1970), *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55, 9th ed., US Government Printing Office.

Ait-Sahalia, Y., and D. Xiu (2017), Using principal component analysis to estimate a high dimensional factor model with high-frequency data, *Journal of Econometrics*, *201*(2), 384–399.

Akaike, H. (1987), Factor analysis and AIC, *Psychometrika*, *52*(3), 317–332, doi: 10.1007/BF02294359.

Alzheimer's Association (2018), 2018 Alzheimer's disease facts and figures, *Alzheimer's & Dementia*, *14*(3), 367–429.

Anderson, G. W., A. Guionnet, and O. Zeitouni (2010), *An introduction to random matrices*, Cambridge University Press.

Anderson, T. (2003), *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Statistics, Wiley.

Apostolova, L. G., et al. (2018), Associations of the top 20 Alzheimer disease risk variants with brain amyloidosis, *JAMA neurology*, *75*(3), 328–341.

Asmussen, S., and P. Glynn (2007), *Stochastic simulation: algorithms and analysis*, Springer, New York, NY, USA.

Asmussen, S., and D. Kroese (2006), Improved algorithms for rare event simulation with heavy tails, *Advances in Applied Probability*, *38*, 545–558.

Bai, J., and K. Li (2012), Statistical analysis of factor models of high dimension, *The Annals of Statistics*, *40*(1), 436–465.

Bai, J., and S. Ng (2002), Determining the number of factors in approximate factor models, *Econometrica*, *70*(1), 191–221.

Bai, Z., and H. Saranadasa (1996), Effect of high dimension: By an example of a two sample problem, *Statistica Sinica*, *6*(2), 311–329.

Bai, Z., D. Jiang, J.-F. Yao, and S. Zheng (2009), Corrections to LRT on large-dimensional covariance matrix by RMT, *The Annals of Statistics*, *37*(6B), 3822–3840.

Bai, Z., D. Jiang, J.-f. Yao, and S. Zheng (2013), Testing linear hypotheses in high-dimensional regressions, *Statistics*, *47*(6), 1207–1223.

Barendse, M., F. Oort, and M. Timmerman (2015), Using exploratory factor analysis to determine the dimensionality of discrete responses, *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 87–101.

Barndorff-Nielsen, O., and P. Hall (1988), On the level-error after Bartlett adjustment of the likelihood ratio statistic, *Biometrika*, *75*(2), 374–378.

Bartholomew, D. J., M. Knott, and I. Moustaki (2011), *Latent variable models and factor analysis: A unified approach*, vol. 904, John Wiley & Sons.

Bartlett, M. S. (1950), Tests of significance in factor analysis, *British Journal of Statistical Psychology*, *3*(2), 77–85.

Basu, S., and W. Pan (2011), Comparison of statistical tests for disease association with rare variants, *Genetic epidemiology*, *35*(7), 606–619.

Bentler, P. M., and K.-H. Yuan (1998), Tests for linear trend in the smallest eigenvalues of the correlation matrix, *Psychometrika*, *63*(2), 131–144.

Bianchi, P., M. Debbah, M. Maïda, and J. Najim (2011), Performance of statistical tests for single-source detection using random matrix theory, *IEEE Transactions on Information theory*, *57*(4), 2400–2419.

Bickel, P. J., and E. Levina (2008), Regularized estimation of large covariance matrices, *The Annals of Statistics*, *36*(1), 199–227.

Billingsley, P. (1995), *Probability and Measure*, 3 ed., John Wiley & Sons.

Blanchet, J., and P. Glynn (2008), Effcient rare event simulation for the maximum of heavy-tailed random walks, *Annals of Applied Probability*, *18*(4), 1351–1378.

Bucklew, J. (2013), *Introduction to rare event simulation*, Springer Science & Business Media.

Button, K. S., J. P. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. Robinson, and M. R. Munafò (2013a), Confidence and precision increase with high statistical power, *Nature Reviews Neuroscience*, *14*(8), 585–585.

Button, K. S., J. P. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. Robinson, and M. R. Munafò (2013b), Power failure: why small sample size undermines the reliability of neuroscience, *Nature Reviews Neuroscience*, *14*(5), 365–376.

Cai, T. T. (2017), Global testing and large-scale multiple testing for high-dimensional covariance structures, *Annual Review of Statistics and Its Application*, *4*(1), 423–446.

Cai, T. T., and T. Jiang (2011), Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices, *The Annals of Statistics*, *39*(3), 1496–1525.

Cai, T. T., and Z. Ma (2013), Optimal hypothesis testing for high dimensional covariance matrices, *Bernoulli*, *19*(5B), 2359–2388.

Cai, T. T., W. Liu, and Y. Xia (2013), Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings, *Journal of the American Statistical Association*, *108*(501), 265–277.

Cai, T. T., W. Liu, and Y. Xia (2014), Two-sample test of high dimensional means under dependence, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(2), 349–372.

Candès, E. J., and P. Sur (2020), The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression, *The Annals of Statistics*.

Casella, G., and R. L. Berger (2002), *Statistical inference*, 2 ed., Thomson Learning.

Cattell, R. (1966), The scree test for the number of factors, *Multivariate Behavioral Research*, *1*(2), 245–276.

Chattopadhyay, A., and T.-P. Lu (2019), Gene-gene interaction: the curse of dimensionality, *Annals of Translational Medicine*, *7*(24).

Chen, S. X., and H. Cui (2006), On Bartlett correction of empirical likelihood in the presence of nuisance parameters, *Biometrika*, *93*(1), 215–220.

Chen, S. X., and Y.-L. Qin (2010), A two-sample test for high-dimensional data with applications to gene-set testing, *The Annals of Statistics*, *38*(2), 808–835.

Chen, S. X., L. Peng, and Y.-L. Qin (2009), Effects of data dimension on empirical likelihood, *Biometrika*, *96*(3), 711–722.

Chen, S. X., L.-X. Zhang, and P.-S. Zhong (2010), Tests for high-dimensional covariance matrices, *Journal of the American Statistical Association*, *105*(490), 810–819, doi:10.1198/jasa.2010.tm09560.

Chen, S. X., J. Li, and P.-S. Zhong (2019a), Two-sample and ANOVA tests for high dimensional means, *The Annals of Statistics*, *47*(3), 1443–1474.

Chen, X. (2018), Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications, *The Annals of Statistics*, *46*(2), 642–678.

Chen, Y., and X. Li (2020), Determining the number of factors in high-dimensional generalised latent factor models, *arXiv preprint arXiv:2010.02326*.

Chen, Y., X. Li, and S. Zhang (2019b), Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis, *Psychometrika*, *84*(1), 124–146.

Chiani, M. (2014), Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy–Widom distribution, *Journal of Multivariate Analysis*, *129*, 69–81.

Cleff, T. (2019), *Applied Statistics and Multivariate Data Analysis for Business and Economics: A Modern Approach Using SPSS, Stata, and Excel*, Springer.

Colantuoni, C., et al. (2011), Temporal dynamics and genetic control of transcription in the human prefrontal cortex, *Nature*, *478*(7370), 519.

Cordeiro, G. M., and F. Cribari-Neto (2014), *An introduction to Bartlett correction and bias reduction*, Springer.

Costello, A. B., and J. Osborne (2005), Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis, *Practical assessment, research, and evaluation*, *10*(1), 7.

Davis, A. (1972), On the ratios of the individual latent roots to the trace of a Wishart matrix, *Journal of Multivariate Analysis*, *2*(4), 440–443.

Demmel, J. W. (1997), *Applied Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

DiCiccio, T., P. Hall, and J. Romano (1991), Empirical likelihood is Bartlett-correctable, *the Annals of Statistics*, *19*(2), 1053–1061.

Dobriban, E. (2020), Permutation methods for factor analysis and PCA, *The Annals of Statistics*.

Donoho, D., and J. Jin (2004), Higher criticism for detecting sparse heterogeneous mixtures, *The Annals of Statistics*, *32*(3), 962–994.

Donoho, D., and J. Jin (2015), Higher criticism for large-scale inference, especially for rare and weak effects, *Statistical Science*, *30*(1), 1–25.

Dumas-Mallet, E., K. S. Button, T. Boraud, F. Gonon, and M. R. Munafò (2017), Low statistical power in biomedical science: a review of three human research domains, *Royal Society Open Science*, *4*(2), 160,254.

Dumitriu, I., and A. Edelman (2002), Matrix models for beta ensembles, *Journal of Mathematical Physics*, *43*(11), 5830–5847.

Dumitriu, I., and A. Edelman (2003), Eigenvalue statistics for beta ensembles, *PhD thesis , Massachusetts Institute of Technology*.

Dupuis, P., K. Leder, and H. Wang (2007), Importance sampling for sums of random variables with regularly varying tails, *ACM Transactions on Modeling and Computer Simulation*, *17*(3), 1–14.

Durrett, R. (2019), *Probability: theory and examples*, 5 ed., Cambridge University Press.

Fabrigar, L. R., and D. T. Wegener (2011), *Exploratory factor analysis*, Oxford University Press.

Fabrigar, L. R., D. T. Wegener, R. C. MacCallum, and E. J. Strahan (1999), Evaluating the use of exploratory factor analysis in psychological research, *Psychological Methods*, *4*(3), 272–299.

Fan, J. (1996), Test of significance based on wavelet thresholding and Neyman's truncation, *Journal of the American Statistical Association*, *91*(434), 674–688.

Fan, J., Y. Liao, and J. Yao (2015), Power enhancement in high-dimensional cross-sectional tests, *Econometrica*, *83*(4), 1497–1541.

Finch, W. H., and M. E. H. Finch (2016), Fitting exploratory factor analysis models with high dimensional psychological data, *Journal of Data Science*, *14*(3), 519–537.

Frahm, G. (2004), Generalized elliptical distributions: theory and applications, Ph.D. thesis, Universität zu Köln.

Gaetan, C., and X. Guyon (2010), *Spatial statistics and modeling*, vol. 90, Springer.

Gamelin, T. (2001), *Complex analysis*, 1 ed., Springer Science & Business Media.

Goeman, J. J., S. A. Van De Geer, and H. C. Van Houwelingen (2006), Testing against a high dimensional alternative, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(3), 477–493.

Gorsuch, R. L. (1988), Exploratory factor analysis, in *Handbook of Multivariate Experimental Psychology*, edited by J. R. Nesselroade and R. B. Cattell, pp. 231–258, Springer US.

Gregory, K. B., R. J. Carroll, V. Baladandayuthapani, and S. N. Lahiri (2015), A two-sample test for equality of means in high dimension, *Journal of the American Statistical Association*, *110*(510), 837–849.

Guyon, X. (1995), *Random fields on a network: modeling, statistics, and applications*, Springer-Verlag, New York.

Hall, P. (1979), On the rate of convergence of normal extremes, *Journal of Applied Probability*, *16*(2), 433–439.

Hall, P., and J. Jin (2010), Innovated higher criticism for detecting sparse signals in correlated noise, *The Annals of Statistics*, *38*(3), 1686–1732.

Harlow, L. L., and F. L. Oswald (2016), Big data in psychology: Introduction to the special issue., *Psychological Methods*, *21*(4), 447.

Hayashi, K., P. Bentler, and K.-H. Yuan (2007), On the likelihood ratio test for the number of factors in exploratory factor analysis, *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 505–526, doi:10.1080/10705510701301891.

He, X., and Q.-M. Shao (2000), On parameters of increasing dimensions, *Journal of Multivariate Analysis*, *73*(1), 120–135.

He, Y., and G. Xu (2018), Estimating tail probabilities of the ratio of the largest eigenvalue to the trace of a Wishart matrix, *Journal of Multivariate Analysis*, *166*, 320–334.

He, Y., T. Jiang, J. Wen, and G. Xu (2021a), Likelihood ratio test in multivariate linear regression: From low to high dimension, *Statistica Sinica*, *31*, 1–24.

He, Y., B. Meng, Z. Zeng, and G. Xu (2021b), On the phase transition of Wilks' phenomenon, *Biometrika*, *accepted*.

He, Y., Z. Wang, and G. Xu (2021c), A note on the likelihood ratio test in high-dimensional exploratory factor analysis, *Psychometrika*.

He, Y., G. Xu, C. Wu, and W. Pan (2021d), Asymptotically independent U-statistics in high-dimensional testing, *Annals of Statistics*, *49*(1), 154–181.

He, Y., G. Xu, C. Wu, and W. Pan (2021e), Supplement to "Asymptotically Independent U-Statistics in High-Dimensional Testing", *Annals of Statistics*, doi: 10.1214/20-AOS1951SUPP.

Heinig, M., et al. (2010), A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk, *Nature*, *467*(7314), 460.

Heyde, C. C., and B. M. Brown (1970), On the departure from normality of a certain class of martingales, *The Annals of Mathematical Statistics*, *41*(6), 2161–2165.

Hjort, N. L., I. W. McKeague, and I. Van Keilegom (2009), Extending the scope of empirical likelihood, *The Annals of Statistics*, *37*(3), 1079–1111.

Ho, H.-C., and T. Hsing (1996), On the asymptotic joint distribution of the sum and maximum of stationary normal random variables, *Journal of Applied Probability*, *33*(1), 138–145.

Ho, H.-C., and W. P. McCormick (1999), Asymptotic distribution of sum and maximum for Gaussian processes, *Journal of Applied Probability*, *36*(4), 1031–1044.

Hoeffding, W. (1948), A class of statistics with asymptotically normal distribution, *The Annals of Mathematical Statistics*, *19*(3), 293–325.

Horn, J. L. (1965), A rationale and test for the number of factors in factor analysis, *Psychometrika*, *30*(2), 179–185.

Hsing, T. (1995), A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables, *The Annals of Probability*, *23*(2), 938–947.

Huang, T., L. Lan, X. Fang, P. An, J. Min, and F. Wang (2015), Promises and challenges of big data computing in health sciences, *Big Data Research*, *2*(1), 2–11.

Ibragimov, I. A., and Y. V. Linnik (1971), *Independent and stationary sequences of random variables*, 443 pp., Wolters-Noordhoff Publishing, Groningen.

James, B., K. James, and Y. Qi (2007), Limit distribution of the sum and maximum from multivariate Gaussian sequences, *Journal of multivariate analysis*, *98*(3), 517–532.

Jansen, I. E., J. E. Savage, K. Watanabe, J. Bryois, D. M. Williams, et al. (2019), Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk, *Nature Genetics*, *51*, 404–413.

Jiang, D., T. Jiang, and F. Yang (2012), Likelihood ratio tests for covariance matrices of high-dimensional normal distributions, *Journal of Statistical Planning and Inference*, *142*(8), 2241–2256.

Jiang, T. (2004), The asymptotic distributions of the largest entries of sample correlation matrices, *The Annals of Applied Probability*, *14*(2), 865–880.

Jiang, T., and Y. Qi (2015), Likelihood ratio tests for high-dimensional normal distributions, *Scandinavian Journal of Statistics*, *42*(4), 988–1009.

Jiang, T., and F. Yang (2013), Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions, *The Annals of Statistics*, *41*(4), 2029–2074.

Jiang, T., K. Leder, and G. Xu (2017), Rare-event analysis for extremal eigenvalues of white Wishart matrices, *The Annals of Statistics*, *45*, 1609–1637.

Johansson, K. (2000), Shape fluctuations and random matrices, *Communications in Mathematical Physics*, *209*(2), 437–476.

Johnstone, I. M. (2001), On the distribution of the largest eigenvalue in principal components analysis, *The Annals of Statistics*, *29*(2), 295–327.

Kaiser, H. F. (1960), The application of electronic computers to factor analysis, *Educational and Psychological Measurement*, *20*(1), 141–151.

Kan, R. (2008), From moments of sum to moments of product, *Journal of Multivariate Analysis*, *99*(3), 542–554.

Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa (2010), KEGG for representation and analysis of molecular networks involving diseases and drugs, *Nucleic Acids Research*, *38*(suppl 1), D355–D360.

Karoui, N. (2006), A rate of convergence result for the largest eigenvalue of complex white wishart matrices, *The Annals of Probability*, *34*(6), 2077–2117.

Keeling, K. B. (2000), A regression equation for determining the dimensionality of data, *Multivariate Behavioral Research*, *35*(4), 457–468.

Kim, J., J. R. Wozniak, B. A. Mueller, X. Shen, and W. Pan (2014), Comparison of statistical tests for group differences in brain functional networks, *NeuroImage*, *101*, 681–694.

Kim, J., Y. Zhang, and W. Pan (2016), Powerful and adaptive testing for multi-trait and multi-SNP associations with GWAS and sequencing data, *Genetics*, *203*(2), 715–731.

Kim, T. Y. (1994), Moment bounds for non-stationary dependent sequences, *Journal of Applied Probability*, *31*(3), 731–742.

Kortun, A., M. Sellathurai, T. Ratnarajah, and C. Zhong (2012), Distribution of the ratio of the largest eigenvalue to the trace of complex Wishart matrices, *IEEE Transactions on Signal Processing*, *60*(10), 5527–5532.

Krzanowski, W. (2000), *Principles of multivariate analysis*, Oxford University Press.

Kuriki, S., and A. Takemura (2001), Tail probabilities of the maxima of multilinear forms and their applications, *The Annals of Statistics*, pp. 328–371.

Lan, W., R. Luo, C.-L. Tsai, H. Wang, and Y. Yang (2015), Testing the diagonality of a large covariance matrix in a regression setting, *Journal of Business & Economic Statistics*, *33*(1), 76–86.

Lawley, D. N., and A. E. Maxwell (1962), Factor analysis as a statistical method, *Journal of the Royal Statistical Society. Series D (The Statistician)*, *12*(3), 209–229.

Ledoit, O., and M. Wolf (2002), Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, *The Annals of Statistics*, *30*(4), 1081–1102.

Lehmann, E. (2012), Statistical hypotheses, *Breakthroughs in Statistics: Foundations and Basic Theory*, p. 67.

Leung, D., and M. Drton (2018), Testing independence in high dimensions with sums of rank correlations, *The Annals of Statistics*, *46*(1), 280–307.

Li, D., and L. Xue (2015), Joint limiting laws for high-dimensional independence tests, *ArXiv e-prints*.

Li, J., and S. X. Chen (2012), Two sample tests for high-dimensional covariance matrices, *The Annals of Statistics*, *40*(2), 908–940.

Liu, J., and G. Xu (2014a), On the conditional distributions and the efficient simulations of exponential integrals of Gaussian random fields, *Annals of Applied Probability*, *24*, 1691–1738.

Liu, J., and G. Xu (2014b), Efficient simulations for the exponential integrals of Hölder continuous Gaussian random fields, *The ACM Transactions on Modeling and Computer Simulation*, *24*, 9:1–9:24.

Liu, W.-D., Z. Lin, and Q.-M. Shao (2008), The asymptotic distribution and Berry-Esseen bound of a new test for independence in high dimension with an application to stochastic optimization, *Ann. Appl. Probab.*, *18*(6), 2337–2366, doi:10.1214/08-AAP527.

Luke, Y. L. (1969), *Special functions and their approximations*, vol. 2, Academic Press.

Luo, L., C. Arizmendi, and K. M. Gates (2019), Exploratory Factor Analysis (EFA) Programs in R, *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(5), 819–826.

Ma, Z. (2012), Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices, *Bernoulli*, *18*(1), 322–359.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, et al. (2009), Finding the missing heritability of complex diseases, *Nature*, *461*(7265), 747.

Maruyama, Y., and T. Seo (2003), Estimation of moment parameter in elliptical distributions, *Journal of the Japan Statistical Society*, *33*(2), 215–229.

Marx, V. (2013), The big challenges of big data, *Nature*, *498*(7453), 255–260.

McCormick, W., and Y. Qi (2000), Asymptotic distribution for the sum and maximum of Gaussian processes, *Journal of Applied Probability*, *37*(4), 958–971.

Mosteller, F., and R. A. Fisher (1948), Questions and answers, *The American Statistician*, *2*(5), 30–31.

Muirhead, R. J. (2009), *Aspects of Multivariate Statistical Theory*, John Wiley & Sons.

Mukherjee, B. N. (1970), Likelihood ratio tests of statistical hypotheses associated with patterned covariance matrices in psychology, *British Journal of Mathematical and Statistical Psychology*, *23*(2), 89–120.

Nadler, B. (2011), On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix, *Journal of Multivariate Analysis*, *102*(2), 363–371.

Neyman, J., and E. S. Pearson (1933), On the problem of the most efficient tests of statistical hypotheses, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *231*(694-706), 289–337.

Owen, A. (1990), Empirical likelihood ratio confidence regions, *The Annals of Statistics*, pp. 90–120.

Paindaveine, D., and G. Van Bever (2014), Inference on the shape of elliptical distributions based on the MCD, *Journal of Multivariate Analysis*, *129*, 125–144.

Palit, S., C. Heuser, G. P. de Almeida, F. J. Theis, and C. E. Zielinski (2019), Meeting the challenges of high-dimensional single-cell data analysis in immunology, *Frontiers in Immunology*, *10*, 1515.

Pan, W., J. Kim, Y. Zhang, X. Shen, and P. Wei (2014), A powerful and adaptive association test for rare variants, *Genetics*, *197*(4), 1081–1095.

Paul, D., and A. Aue (2014), Random matrix theory in statistics: A review, *Journal of Statistical Planning and Inference*, *150*, 1–29.

Péché, S. (2009), Universality results for the largest eigenvalues of some sample covariance matrix ensembles, *Probability Theory and Related Fields*, *143*(3-4), 481–516.

Peng, Z., and S. Nadarajah (2003), On the joint limiting distribution of sums and maxima of stationary normal sequence, *Theory of Probability & Its Applications*, *47*(4), 706–709.

Pham, T. D., and L. T. Tran (1985), Some mixing properties of time series models, *Stochastic Processes and their Applications*, *19*(2), 297–303.

Pituch, K. A., and J. P. Stevens (2015), *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS*, Routledge.

Portnoy, S. (1985), Asymptotic behavior of M estimators of p regression parameters when p^2/n is large; II. Normal approximation, *The Annals of Statistics*, pp. 1403–1417.

Portnoy, S. (1988), Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity, *The Annals of Statistics*, pp. 356–366.

Preacher, K. J., and R. C. MacCallum (2002), Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes, *Behavior genetics*, *32*(2), 153–161.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1992), *Numerical recipes in Fortran 77: the art of scientific computing*, Cambridge University Press.

Prince, M., R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri (2013), The global prevalence of dementia: a systematic review and metaanalysis, *Alzheimer's & Dementia*, *9*(1), 63–75.

Qin, Y., H. K. Yalamanchili, J. Qin, B. Yan, and J. Wang (2015), The current status and challenges in computational analysis of genomic big data, *Big Data Research*, *2*(1), 12–18.

Schott, J. R. (2007), A test for the equality of covariance matrices when the dimension is large relative to the sample sizes, *Computational Statistics & Data Analysis*, *51*(12), 6535–6542.

Schuurmann, F., P. Krishnaiah, and A. Chattopadhyay (1973), On the distributions of the ratios of the extreme roots to the trace of the Wishart matrix, *Journal of Multivariate Analysis*, *3*(4), 445–453.

Schwarz, G. (1978), Estimating the dimension of a model, *The Annals of Statistics*, *6*(2), 461–464.

Sham, P. C., and S. M. Purcell (2014), Statistical power and significance testing in large-scale genetic studies, *Nature Reviews Genetics*, *15*(5), 335–346.

Shao, Q.-M., and W.-X. Zhou (2014), Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices, *The Annals of Probability*, *42*(2), 623–648.

Siegmund, D. (1976), Importance sampling in the Monte Carlo study of sequential tests, *The Annals of Statistics*, *4*, 673–684.

Soshnikov, A. (2002), A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices, *Journal of Statistical Physics*, *108*(5-6), 1033–1056.

Srivastava, M. S., and M. Du (2008), A test for the mean vector with fewer observations than the dimension, *Journal of Multivariate Analysis*, *99*(3), 386–402.

Srivastava, M. S., and H. Yanagihara (2010), Testing the equality of several covariance matrices with fewer observations than the dimension, *Journal of Multivariate Analysis*, *101*(6), 1319–1329.

Srivastava, R., P. Li, and D. Ruppert (2016), RAPTT: An exact two-sample test in high dimensions using random projections, *Journal of Computational and Graphical Statistics*, *25*(3), 954–970.

Steiger, J. H. (2016), Notes on the Steiger–Lind (1980) handout, *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(6), 777–781.

Sundberg, R., and U. Feldmann (2016), Exploratory factor analysis – Parameter estimation and scores prediction with high-dimensional data, *Journal of Multivariate Analysis*, *148*, 49–59.

Sur, P., and E. J. Candès (2019), A modern maximum-likelihood theory for high-dimensional logistic regression, *Proceedings of the National Academy of Sciences*, *116*(29), 14,516–14,525.

Sur, P., Y. Chen, and E. J. Candès (2019), The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square, *Probability Theory and Related Fields*, *175*(1-2), 487–558.

Taliun, D., et al. (2021), Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program, *Nature*, *590*(7845), 290–299.

Tang, C. Y., and C. Leng (2010), Penalized high-dimensional empirical likelihood, *Biometrika*, *97*(4), 905–920.

Thompson, B. (2004), *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*, American Psychological Association.

Tse, D., and P. Viswanath (2005), *Fundamentals of wireless communication*, Cambridge University Press.

Tucker, L. R., and C. Lewis (1973), A reliability coefficient for maximum likelihood factor analysis, *Psychometrika*, *38*(1), 1–10.

Ushakov, N. G. (2011), *Selected topics in characteristic functions*, Walter de Gruyter.

Van der Vaart, A. W. (2000), *Asymptotic statistics*, vol. 3, Cambridge University Press.

Vershynin, R. (2018), *High-dimensional probability: An introduction with applications in data science*, vol. 47, Cambridge University Press.

Wang, C., M.-H. Chen, E. Schifano, J. Wu, and J. Yan (2016), Statistical methods and computing for big data, *Statistics and Its Interface*, *9*(4), 399.

Wang, L. (2011), GEE analysis of clustered binary data with diverging number of covariates, *The Annals of Statistics*, *39*(1), 389–417.

Wang, L., P. Jia, R. D. Wolfinger, X. Chen, and Z. Zhao (2011), Gene set analysis of genome-wide association studies: methodological issues and perspectives, *Genomics*, *98*(1), 1–8.

Wei, L., O. Tirkkonen, P. Dharmawansa, and M. McKay (2012), On the exact distribution of the scaled largest eigenvalue, in *2012 IEEE International Conference on Communications (ICC)*, pp. 2422–2426, IEEE.

Whittaker, E. T., and G. N. Watson (1996), *A course of modern analysis*, Cambridge University Press.

Wilks, S. S. (1932), Certain generalizations in the analysis of variance, *Biometrika*, pp. 471–494.

Wu, C., G. Xu, and W. Pan (2019), An adaptive test on high-dimensional parameters in generalized linear models, *Statistica Sinica*, *29*(4), 2163–2186.

Xia, Y., T. Cai, and T. T. Cai (2015), Testing differential networks with applications to the detection of gene-gene interactions, *Biometrika*, *102*(2), 247–266.

Xu, G., G. Lin, and J. Liu (2014), Rare-event simulation for the stochastic Korteweg–de Vries equation, *SIAM/ASA Journal on Uncertainty Quantification*, *2*(1), 698–716.

Xu, G., L. Lin, P. Wei, and W. Pan (2016), An adaptive two-sample test for high-dimensional means, *Biometrika*, *103*(3), 609–624.

Xu, J., S. L. Murphy, K. D. Kochanek, B. Bastian, and E. Arias (2018), Deaths: Final data for 2016, *National Vital Statistics Reports*, *67*(5).

Xu, Y., and P. Wang (2020), Rare event estimation of high dimensional problems with confidence intervals, in *Proceedings of the IISE Annual Conference*, pp. 31A–36A, Institute of Industrial and Systems Engineers (IISE).

Xu, Z., G. Xu, and W. Pan (2017), Adaptive testing for association between two random vectors in moderate to high dimensions, *Genetic epidemiology*, *41*(7), 599–609.

Yang, Q., and G. Pan (2017), Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices, *Journal of the American Statistical Association*, *112*(517), 188–200.

Yu, K., Q. Li, A. W. Bergen, R. M. Pfeiffer, P. S. Rosenberg, N. Caporaso, P. Kraft, and N. Chatterjee (2009), Pathway analysis by adaptive combination of p-values, *Genetic epidemiology*, *33*(8), 700–709.

Zheng, S. (2012), Central limit theorems for linear spectral statistics of large dimensional F-matrices, *Annales de l'IHP Probabilités et statistiques*, *48*(2), 444–476.

Zhong, P.-S., and S. X. Chen (2011), Tests for high-dimensional regression coefficients with factorial designs, *Journal of the American Statistical Association*, *106*(493), 260–274.

Zwillinger, D. (2002), *CRC standard mathematical tables and formulae*, 31st ed., CRC Press.